AD_____

Award Number: DAMD17-01-1-0328

TITLE: Computer-Aided Characterization of Breast Masses on Volumetric Ultrasound Images: An Adjunct to Mammography

PRINCIPAL INVESTIGATOR: Berkman Sahiner, M.D., Ph.D.

CONTRACTING ORGANIZATION: University of Michigan
Ann Arbor, Michigan 48109-1274

REPORT DATE: October 2003

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 2003	Annual (6 Sep 2002 - 5 Sep 2003)

4. TITLE AND SUBTITLE
Computer-Aided Characterization of Breast Masses on volumetric Ultrasound Images: An Adjunct to Mammography

5. FUNDING NUMBERS
DAMD17-01-1-0328

6. AUTHOR(S)
Berkman Sahiner, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Michigan
Ann Arbor, Michigan 48109-1274

E-Mail: berki@umich.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

The purpose of this project is to develop computer vision techniques for the analysis of sonographic images of breast masses, and to combine computerized sonographic and mammographic analyses. The techniques developed in this project are aimed at providing a second opinion to the radiologists in the task of making a biopsy recommendation. In the second year of the project, we have (1)compared the accuracy of the classifier designed in the first year of this project to that of experienced radiologists; (2) conducted studies o the effect of the developed classifier on radiologists' characterization of breast masses on ultrasound images; and (3) investigated methods for combining computer classification methods based on ultrasound and mammogram images. Our results indicate that the accuracy of our computer classifier is similar to that of experienced breast radiologists on our data set. We have also shown that experienced radiologists can significantly ($p<0.006$) improve their mass characterization accuracy on sonograms when aided by our algorithm. Our results on combining computer classification methods based on ultrasound and mammogram images indicate that multi-modality computer-aided diagnosis may further improve the classification accuracy.

14. SUBJECT TERMS
Computer-aided diagnosis; ultrasonography; breast masses; breast cancer detection;

15. NUMBER OF PAGES
62

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

(4) Introduction

At present, biopsy is the gold standard in breast lesion characterization. However, the positive breast biopsy rate is only about 15-30%. This means that 70-85% of breast biopsies are performed for benign lesions. In order to reduce patient anxiety and morbidity, as well as to decrease health care costs, it is desirable to reduce the number of benign biopsies without missing malignancies. Mammography and sonography are two low-cost imaging modalities that may be improved so that radiologists can obtain more accurate diagnostic information to differentiate malignant and benign lesions. Computerized analysis of the lesions on these images is one of the promising tools that may improve the radiologists' accuracy in characterizing these lesions by providing a consistent and reliable second opinion to radiologists.

In this project, our goal is to analyze volumetric images to improve the accuracy of computerized sonographic breast lesion characterization, and to combine these characterization results with those obtained by computerized analysis of mammograms. Computerized image analysis, feature extraction, and classification methods will be developed to characterize breast masses on three-dimensional or volumetric ultrasound images. The output of the classifier will be a computer rating related to the likelihood of malignancy of the mass. The accuracy of this rating will be studied by comparing it to the biopsy results. We will then combine this rating with a similar rating obtained by computerized analysis of the mammograms of the same patient. The combined classifier is expected to be more accurate than either classifier alone.

(5) Body

In the current project year (9/6/02-9/5/03), we have performed the following studies:

(A) Collection of Database

(a) Volumetric sonograms: Up to the end of the second year of the project, have collected 134 volumetric scans from 92 patients. We are working hard on reminding the radiologists to save these images and providing them feedback on the quality of the data that has been collected.

(b) 2D sonograms: Since February 2002, we have collected 795 2D sonograms from the 92 patients included in our volumetric sonogram data set.

(c) Build a database of mammograms of the corresponding mammograms: In the second year of the project, we have started collecting a database of mammograms corresponding to the 3D sonogram sets. From patient folders, we have selected mammograms that are closest in time to the acquisition of ultrasound images in our data set, and also prior to biopsy. Up to this point, we have digitized mammograms of more than 60 patients in our 3D data set.

(d) Identify and rank the lesions by radiologists: For the ultrasound data set, we have developed a graphical user interface (GUI) to facilitate the identification and rating of the lesions. An example of the GUI is shown in Fig. 1. The biopsied mass in each volume was identified by the expert radiologist, Dr. Marilyn Roubidoux, using clinical US and mammographic images to confirm that the 3D images contained the clinically suspicious mass. Four other radiologists ranked the lesions in terms of their shape, margins, echogenecity, through transmission, other findings (such as ductal extension and calcifications), likelihood of malignancy, and the overall impression. In order to ensure that all radiologists ranked the same lesion, an arrow pointing to the identified lesion was provided. The radiologists could navigate through the slices of the 3D set, adjust brightness and contrast, and view the images as a cine-loop with a desired delay between slices.

For the mammography data set, the expert mammographer in this project, Dr. Mark Helvie, has read cases from 60 patients, which involves the identification of the biopsied lesion, and its rating for malignancy and visibility.

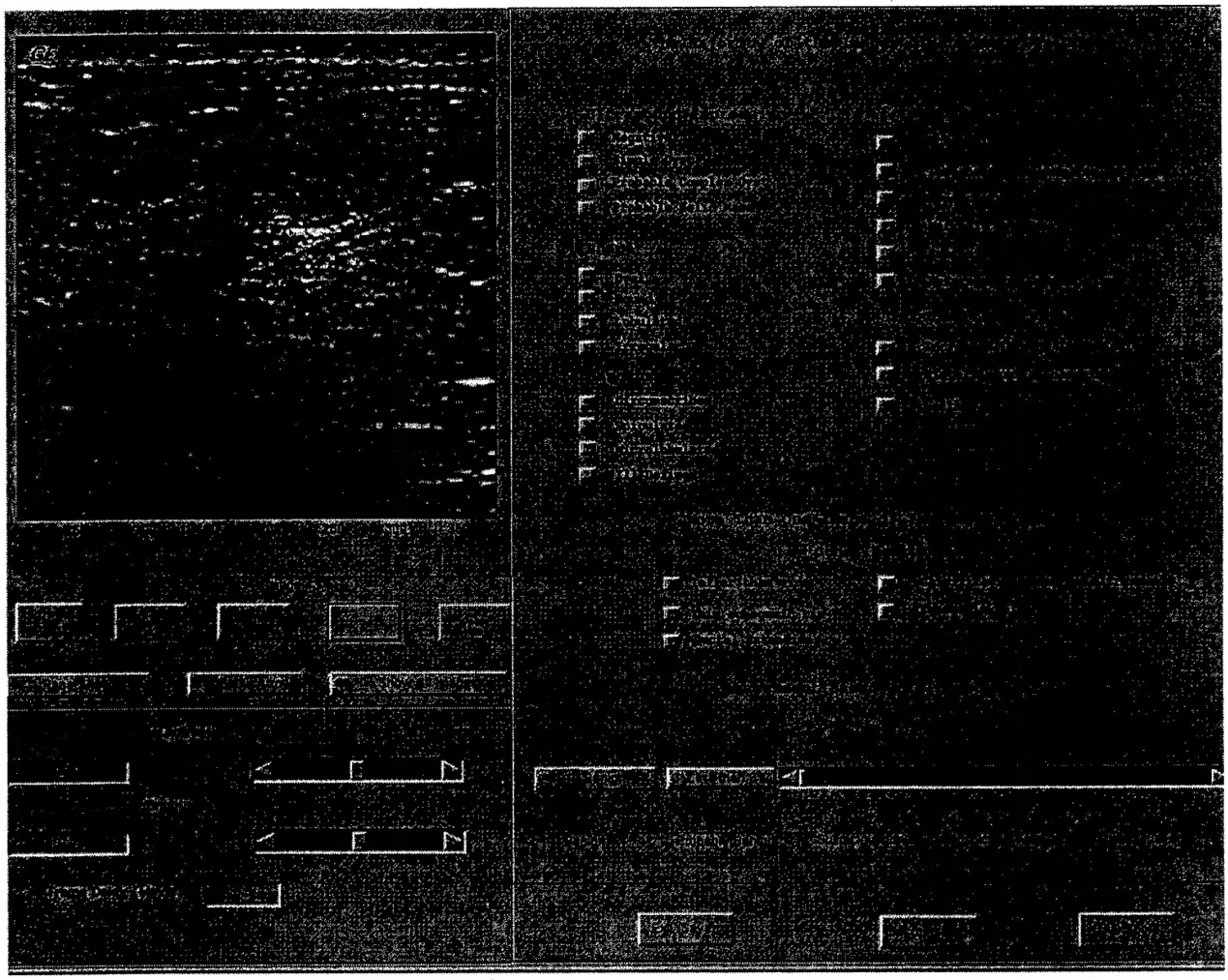

Fig. 1. The GUI developed for the identifying and rating the ultrasound lesions. The same GUI was used by the radiologists in the first part of the ROC experiment described in Section (C) for ranking the likelihood of malignancy of the lesions without CAD.

(B) Evaluation of Computer Classification Accuracy on the 3D Ultrasound Data Set

As explained in our first yearly report, we developed computerized image segmentation, feature extraction, and classification methods for the 3D ultrasound data in the first year of the

project. In the second year of the project, we evaluated the developed computerized mass characterization method by (i) comparing its accuracy to that of experienced radiologists, and (ii) investigating its robustness with respect to the accuracy of the segmentation method. In the rest of this section, we summarize these two evaluation studies. For details on these studies, please refer to Appendix 5.

Our data set for the evaluation consisted of 3D ultrasound images from 102 who had a solid mass deemed suspicious or highly suggestive of malignancy. All patients underwent biopsy or fine needle aspiration as part of their clinical management. Fifty-eight masses were malignant and 44 were benign.

In the first evaluation study, we compared the accuracy of the designed computerized mass characterization method to that of four experienced breast radiologists (RAD1-RAD4), who were either fellowship-trained in breast imaging or had over 25 years of experience in breast imaging. All four radiologists were MQSA qualified and had mammographic and US interpretation experience ranging from 2 to 25 years (mean: 11.3 years). The location of the center of mass was displayed on each slice so that all the radiologists would rank the same mass if more than one mass existed in the volume. There was no time limitation for the radiologists to read a case. The case reading order was randomized for each radiologist. The malignancy rating was entered by means of a slide bar. Before participating in the study, the radiologists were trained on five cases that were not part of the data set of 102 cases. The classification accuracy was compared using the area under the receiver operating characteristic (ROC) curve, A_z, as well as the partial area index, $A_z^{(0.9)}$. $A_z^{(0.9)}$ is defined as the area under the ROC curve above a sensitivity threshold of 0.9 ($TPF_0 = 0.9$) normalized to the total area above TPF_0, which is equal to ($1-TPF_0$) [1].

The ROC curves for radiologists' malignancy ratings are shown in Fig. 2. The computer and radiologist A_z values and $A_z^{(0.9)}$ values are compared in Table 1. The area A_z under the ROC curve

for radiologists RAD1-RAD4 varied between 0.84±0.04 and 0.92±0.03, which are lower than or equal to that of the 3D computer classifier. The average A_z value, obtained by averaging the slope and intercept parameters (a and b in ROC analysis) of the individual ROC curves was 0.87. The difference between the A_z values of the individual radiologists and the computer did not reach statistical significance ($p>0.05$). The $A_z^{(0.9)}$ value of the computer classifier was higher than the $A_z^{(0.9)}$ values of all four radiologists, and achieved statistical significance for three of the four ($p=0.02$, 0.04, and 0.008 for RAD1, RAD2, and RAD4, respectively).

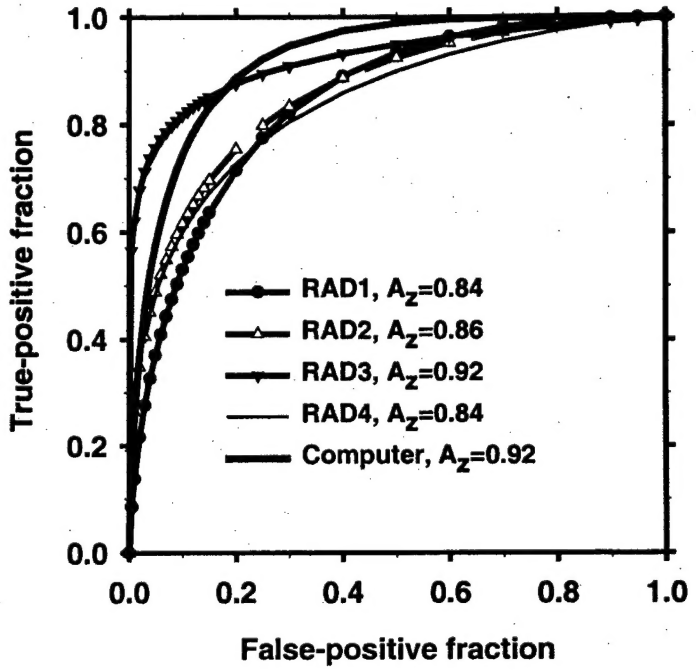

Fig. 2. ROC curves for the computer and for the four radiologists who participated in the malignancy rating experiment. The difference between the computer's A_z value and that of any of the four radiologists did not achieve statistical significance. However, the computer classifier was significantly ($p<0.05$) more accurate than three of the four radiologists at high sensitivity (TPF>0.9).

	A_z	$A_z^{(0.9)}$
Computer classifier	0.92±0.03	0.65±0.09
RAD1	0.84±0.04	0.41±0.10*
RAD2	0.86±0.04	0.37±0.11*
RAD3	0.92±0.03	0.44±0.14
RAD4	0.84±0.04	0.28±0.11*

Table 1: The area under the ROC curve (Az), and the area under the ROC curve above a sensitivity threshold of 0.9 ($A_z^{(0.9)}$) for the computer classifier and the four radiologists. The radiologists' results that are significantly ($p<0.05$) different from the computer results are noted with an asterisk.

In the second evaluation study, we assessed the robustness of our computerized characterization method to the initialization of our segmentation technique. The active contour segmentation method that we developed in the first year of the project requires an initial boundary to start iterating towards the optimal contour. In our method, the initial boundary was defined by a 3D ellipsoid that approximated the mass shape. The ellipsoid was placed in the volume by one of the radiologists (RAD1) using an interactive graphical user interface (GUI). The robustness of the 3D segmentation method to active contour initialization was studied by translating, rotating, and scaling the 3D ellipsoid. There are many possibilities as to how these three operations (moving, rotating, and scaling) can be combined to modify the initial ellipsoid. In Table 2, the classification results are presented when these three operations are performed one at a time. Row 1 shows the A_z value when the original ellipsoid is used. The ellipsoid was scaled in rows 2-3, translated in rows 4-6, and rotated in row 7. For the magnitudes of scaling, translation and rotation studied in Table 2, the variation of the A_z value was within two standard deviations of A_z value provided by the LABROC program [2]. We thus concluded that although the classification

accuracy depends on the initialization, the A_z value did not significantly deteriorate when the initial contour was scaled, rotated, or translated by a moderate amount.

Scale	Rotation (degrees)	x-translation (pixels)	y-translation (pixels)	A_z
1	0	0	0	0.92±0.03
1.3	0	0	0	0.87±0.04
0.8	0	0	0	0.89±0.03
1	0	10	10	0.89±0.03
1	0	10	-10	0.85±0.04
1	0	-10	10	0.88±0.03
1	0	-10	-10	0.87±0.04
1	15	0	0	0.92±0.03

Table 2: The dependence of the computer classification accuracy on the variation of the initial contour. The effects of three transformation parameters, namely, scaling, translation and rotation of the initial ellipsoid, was investigated by moving the initial ellipsoid using one of these three parameters at a time. A translation by ±10 pixels in the image plane corresponded to approximately ±1 mm.

(C) The Effect of Computer-Aided Diagnosis on Radiologists' Characterization Accuracy

We conducted an ROC study to investigate if the classifier developed in the first year of the project would improve radiologists' accuracy in differentiation of malignant and benign breast masses on ultrasound images. The data set for the ROC study consisted of the 102 ultrasound volumes described in Section B. To ensure that all radiologists rated the biopsied lesion, the location of the lesion was identified with an arrow. The radiologists were free to eliminate he arrow once they recognized the lesion. Five MQSA radiologists participated as observers in the ROC study. Each radiologist read the cases first without CAD, immediately

followed by reading with CAD. When reading without CAD, a radiologist analyzed the mass in the 3D volume using the graphical interface shown in Fig. 1, and provided an estimate of the likelihood of malignancy. Subsequently, the radiologist was presented the computer malignancy score using the second page of the GUI as shown in Fig. 3, and had an option to revise his/her malignancy estimate. In order to enable the radiologists to calibrate the computer scores, two Gaussian curves were fitted to the score distributions for the malignant and benign cases in the data set, were displayed by the GUI. The reading order of the ultrasound volumes was randomized for each observer. The classification accuracy was quantified by using the area under ROC curve, Az. The statistical analysis was performed using the MRMC paradigm developed by Dorfman et al [3].

The computer classifier achieved an Az value of 0.92. The radiologists had an average Az of 0.83 (range: 0.81 to 0.87) without CAD. The accuracy of every radiologist was improved when they read with CAD (range: 0.04 to 0.11), and the average Az of the four radiologists improved to 0.89 (range: 0.85 to 0.93). The improvement was statistically significant (p=0.006). The Az values of the radiologists without and with CAD are summarized in Table 3.

Radiologist	Without CAD	With CAD
1	0.84±0.04	0.89±0.03
2	0.81±0.04	0.85±0.04
3	0.87±0.03	0.91±0.03
4	0.82±0.04	0.93±0.02
5	0.83±0.04	0.90±0.03
Average	0.83	0.89

Table 3: The Az values of the radiologists without and with CAD. The radiologists were statistically significantly more accurate when the read with CAD (p=0.006).

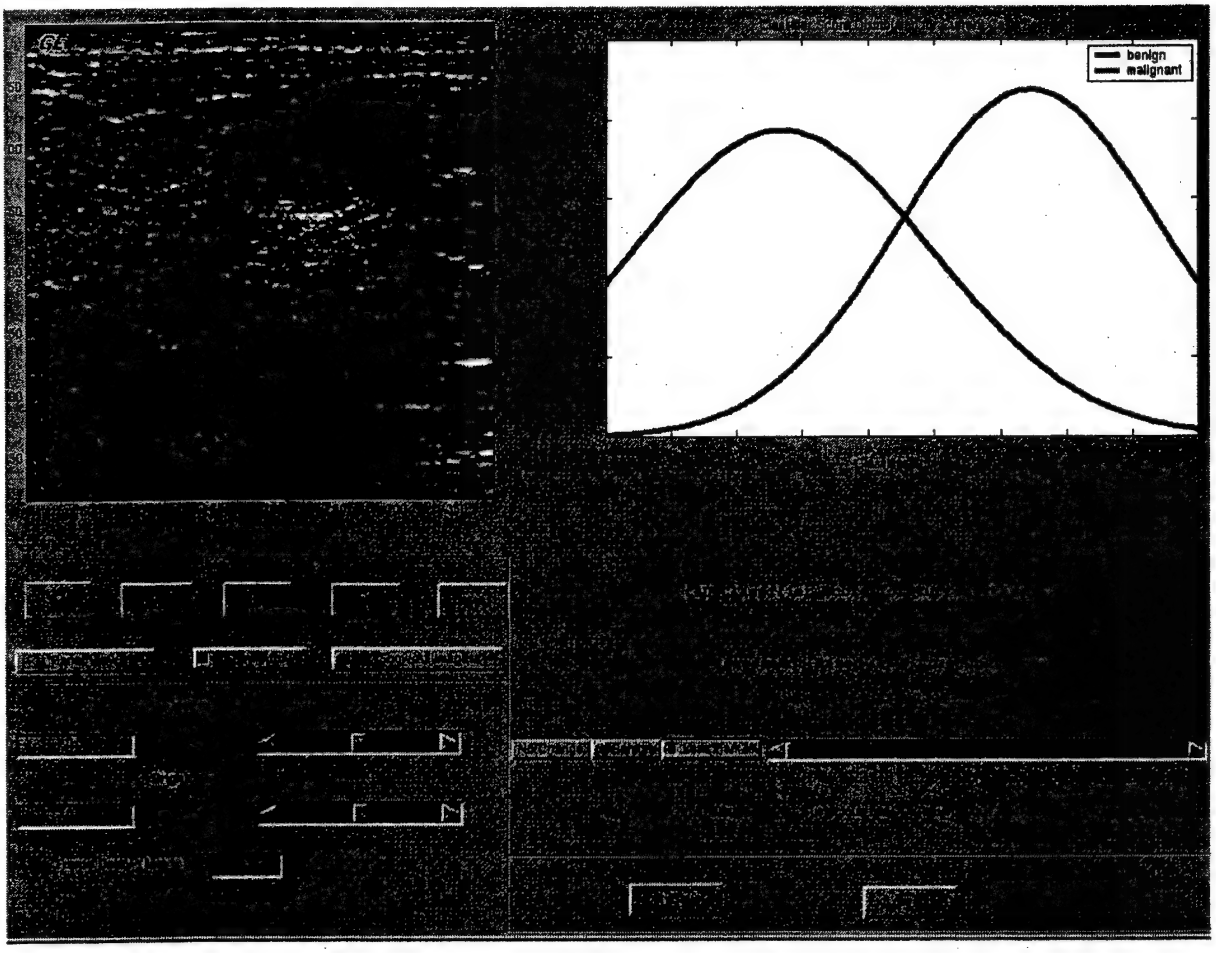

Fig. 3. The GUI page displayed for the radiologists when they ranked the likelihood of malignancy of the lesions with CAD.

(D) Combination of Computerized Classification Techniques Based on Mammograms and 3D Ultrasound Volumes for Improved Accuracy

The data set used in this study consisted of US and mammogram images of biopsy-proven solid breast masses from 60 patients. Thirty of the masses were malignant and 30 were benign. The mammogram set of each mass included between one and three views. The steps in the computerized mammographic mass characterization method included automated mass segmentation, extraction of spiculation features, extraction of morphological features, performing the rubber-band straightening transform, extraction of run-length features, and design of a linear

discriminant classifier [4, 5]. We investigated three methods (A, B and C) for combining the features or scores from different mammographic views and the US volumes. Method A consisted of (i) using available mammographic views of a patient to design a view-based mammographic classifier, (ii) combining view-based mammographic scores into a case-based mammographic score, and (iii) combining cased-based scores from US and mammography. In methods B and C, we first combined the features from different mammographic views of the same patient into a case-based feature for the patient. In method B, case-based features from US and mammograms were combined into a malignancy score using a single classifier. In method C, case-based mammographic features were used to obtain a case-based malignancy score that was combined with the US score as in method A. The leave-one-case-out method was used to obtain test scores for the classifiers. Classifier scores were analyzed using the ROC methodology and the area A_z under the ROC curve.

The A_z values of the classifiers based on US alone and mammography alone were 0.88±0.04 and 0.84±0.05, respectively. In comparison, the average A_z value of four MQSA radiologists in characterizing the same 3D US data was 0.88. The A_z values of methods A, B, and C were 0.91±0.04, 0.90±0.04, and 0.90±0.04, respectively. Using the scores obtained by method A, 40% of the benign masses could be correctly identified without missing a malignancy. However, none of the differences in the A_z values between any two of the classifiers reached statistical significance, possibly due to the limited size of the data set.

(6) Key Research Accomplishments

- Collect databases of 2D and volumetric ultrasound images and a database of corresponding mammograms (Tasks 1b and 1c)

- Identify and rank the lesions for their visibility and likelihood of malignancy by radiologists (Task 1d).
- Compare computerized mass classification methods with radiologists' classification (Task 3)
- Evaluate the effect of the developed classifier on radiologists' characterization of breast masses on ultrasound images (Task 3).
- Perform preliminary studies for combining the mammographic and sonographic computerized mass characterization methods (Task 4).

(7) Reportable Outcomes

In the second year of the project, we have submitted three conference abstracts on computer-aided characterization of breast masses on ultrasound images, all of which have been accepted for publication. Additionally, we have published one conference proceeding paper, and have submitted one journal manuscript on the same topic. We are in the process of writing a manuscript for journal submission on the effect of the developed classifier on radiologists' characterization of breast masses on ultrasound images.

Conference Abstracts:

Sahiner B, Chan HP, Hadjiiski LM, Roubidoux MA, Paramagul C, Helvie MA, Zhou C, "Multi-modality CAD: Combination of computerized classification techniques based on mammograms and 3D ultrasound volumes for improved accuracy in breast mass characterization," to be presented at the *2004 SPIE Medical Imaging Conference*, San Diego, CA, Feb. 14-19, 2004.

Sahiner B, Chan HP, Roubidoux MA, Helvie MA, Bailey J, Hadjiiski LM, "An ROC study on characterization of malignant and benign breast masses in 3D ultrasound volumes: The effect of

computer-aided diagnosis on radiologists' characterization accuracy," to be presented at the *89th Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Chicago, IL, Nov. 30-Dec 5, 2003.

Sahiner B, Chan HP, Paramagul C, Nees AV, Blane CE, Ramachandran A, "Development of a computer classifier for computer-aided characterization of breast masses in 3D ultrasound volumes," to be presented at the *89th Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Chicago, IL, Nov. 30-Dec 5, 2003.

Conference Proceedings:

Sahiner B, Ramachandran A, Chan HP, Hadjiiski LM, Roubidoux MA, Helvie MA, Paramagul C, Nees A, Blane C, Petrick N, Zhou C, "Three-dimensional active contour model for characterization of solid breast masses on three-dimensional ultrasound images," *Proc. SPIE Medical Imaging,* 2003, 5032: 405-413.

Journal Publications:

Sahiner B, Chan HP, Roubidoux MA, Helvie MA, Hadjiiski LM, Ramachandran A, LeCarpentier GL, Nees A, Paramagul C, Blane C, "Computerized characterization of breast masses on 3-D ultrasound volumes," *Med Phys, (submitted)* 2003.

(8) Conclusions

As a result of the support by the USAMRMC BCRP grant, in the second year of this project, we have (1) compared the accuracy of the developed computer classifier to that of experienced radiologists; (2)

conducted studies on the effect of the developed classifier on radiologists' characterization of breast masses on ultrasound images; and (3) investigated methods for combining computer classification methods based on ultrasound and mammogram images. We have also continued database collection, and we are aware that we need to do our utmost to enlarge our data set in the third year of the project.

The results obtained so far are encouraging. We have shown that the accuracy of the classifier designed in the first year of this project is similar to that of experienced breast radiologists on our data set. We have also shown that experienced radiologists can significantly improve their mass characterization accuracy on sonograms when they are aided by our computer algorithm. The comparison of the classifiers based on US alone and mammography alone agrees with the clinical experience that US can be more accurate than mammography for characterization of masses. The results in Section (D) indicate that combining the two modalities can further improve the classification accuracy, although the improvement so far has not achieved statistical significance. Further improvement of the 3D ultrasound characterization methods and improved methods for combination with mammographic computer image analyses can provide radiologists with a powerful aid for decision making, which may help reduce unnecessary biopsies and improve patient care

(9) References

[1] Y. Jiang, C. E. Metz, and R. M. Nishikawa, "A receiver operating characteristic partial area index for highly sensitive diagnostic tests," *Radiology*, vol. 201, pp. 745-750, 1996.

[2] C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," *Statistics in Medicine*, vol. 17, pp. 1033-1053, 1998.

[3] D. D. Dorfman, K. S. Berbaum, and C. E. Metz, "ROC rating analysis: Generalization to the population of readers and cases with the jackknife method," *Invest. Radiol.*, vol. 27, pp. 723-731, 1992.

[4] B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis," *Medical Physics*, vol. 25, pp. 516-526, 1998.

[5] B. Sahiner, H.-P. Chan, N. Petrick, M. A. Helvie, and L. M. Hadjiiski, "Improvement of mammographic mass characterization using spiculation measures and morphological features," *Medical Physics*, vol. 28, pp. 1455-1465, 2001.

(10) Appendix

Copies of the following publications are enclosed with this report:

(1) Sahiner B, Chan HP, Hadjiiski LM, Roubidoux MA, Paramagul C, Helvie MA, Zhou C, "Multi-modality CAD: Combination of computerized classification techniques based on mammograms and 3D ultrasound volumes for improved accuracy in breast mass characterization," to be presented at the *2004 SPIE Medical Imaging Conference*, San Diego, CA, Feb. 14-19, 2004.

(2) Sahiner B, Chan HP, Roubidoux MA, Helvie MA, Bailey J, Hadjiiski LM, "An ROC study on characterization of malignant and benign breast masses in 3D ultrasound volumes: The effect of computer-aided diagnosis on radiologists' characterization accuracy," to be presented at the *89th Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Chicago, IL, Nov. 30-Dec 5, 2003.

(3) Sahiner B, Chan HP, Paramagul C, Nees AV, Blane CE, Ramachandran A, "Development of a computer classifier for computer-aided characterization of breast masses in 3D ultrasound volumes," to be presented at the *89th Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Chicago, IL, Nov. 30-Dec 5, 2003.

(4) Sahiner B, Ramachandran A, Chan HP, Hadjiiski LM, Roubidoux MA, Helvie MA, Paramagul C, Nees A, Blane C, Petrick N, Zhou C, "Three-dimensional active contour model

for characterization of solid breast masses on three-dimensional ultrasound images," *Proc. SPIE Medical Imaging,* 2003, 5032: 405-413.

(5) Sahiner B, Chan HP, Roubidoux MA, Helvie MA, Hadjiiski LM, Ramachandran A, LeCarpentier GL, Nees A, Paramagul C, Blane C, "Computerized characterization of breast masses on 3-D ultrasound volumes," *Med Phys, (submitted)* 2003.

APPENDIX 1

Multi-modality CAD: Combination of computerized classification techniques based on mammograms and 3D ultrasound volumes for improved accuracy in breast mass characterization

Berkman Sahiner, Heang-Ping Chan, Lubomir M. Hadjiiski, Marilyn A. Roubidoux,
Chintana Paramagul, Mark A. Helvie, Chuan Zhou
The University of Michigan Health System, Ann Arbor, MI 48109-0904

PURPOSE: Mammography and ultrasound (US) are two low-cost modalities that are commonly used by radiologists for evaluating breast masses and making biopsy recommendations. The potential of computer-aided diagnosis (CAD) has recently been investigated for both modalities. In order to improve the accuracy of computerized breast mass characterization, we investigated methods for combining information extracted from these two modalities using a data set of corresponding 3D US images and mammograms from the same patients.

METHODS: Our data set consisted of US and mammogram images of biopsy-proven solid breast masses from 60 patients. Thirty of the masses were malignant and 30 were benign. The mammogram set of each mass included between one and three views. The US image volumes were obtained by using an experimental 3D image acquisition system that consisted of a commercially available ultrasound scanner and a mechanical transducer guiding system. The steps in the computerized US mass characterization method included automated 3D mass segmentation, extraction of texture features related to margin characteristics of the mass, extraction of features related to the shape and attenuation characteristics of the mass, and design of a linear discriminant classifier. The steps in the computerized mammographic mass characterization method included automated mass segmentation, extraction of spiculation features, extraction of morphological features, performing the rubber-band straightening transform, extraction of run-length features, and design of a linear discriminant classifier. We investigated three methods (A, B and C) for combining the features or scores from different mammographic views and the US volumes. Method A consisted of (i) using available mammographic views of a patient to design a view-based mammographic classifier, (ii) combining view-based mammographic scores into a case-based mammographic score, and (iii) combining cased-based scores from US and mammography. In methods B and C, we first combined the features from different mammographic views of the same patient into a case-based feature for the patient. In method B, case-based features from US and mammograms were combined into a malignancy score using a single classifier. In method C, case-based mammographic features were used to obtain a case-based malignancy score that was combined with the US score as in method A. The leave-one-case-out method was used to obtain test scores for the classifiers. Classifier scores were analyzed using the Receiver Operating Characteristic (ROC) methodology and the area A_z under the ROC curve.

RESULTS: The A_z values of the classifiers based on US alone and mammography alone were 0.88±0.04 and 0.84±0.05, respectively. In comparison, the average A_z value of four MQSA radiologists in characterizing the same 3D US data was 0.88. The A_z values of methods A, B, and C were 0.91±0.04, 0.90±0.04, and 0.90±0.04, respectively. Using the scores obtained by method A, 40% of the benign masses could be correctly identified without missing a malignancy. However, none of the difference in the A_z values between any two of the classifiers reached statistical significance, possibly due to the limited size of the data set.

NEW OR BREAKTHROUGH WORK: Although computerized breast mass classification methods have been previously developed separately for mammography and for US, no studies to date have investigated computerized combination of information from 3D US images and mammograms. Our study proposes and compares different methods for combining these two modalities for improved accuracy.

CONCLUSION: The comparison of the classifiers based on US alone and mammography alone agrees with the clinical experience that US can be more accurate than mammography for characterization of masses. Our results indicate that combining the two modalities can further improve the classification accuracy.

APPENDIX 2

AN ROC STUDY ON CHARACTERIZATION OF MALIGNANT AND BENIGN BREAST MASSES IN 3D ULTRASOUND VOLUMES: THE EFFECT OF COMPUTER-AIDED DIAGNOSIS ON RADIOLOGISTS' CHARACTERIZATION ACCURACY

Berkman Sahiner, Heang-Ping Chan, Marilyn A Roubidoux, Mark A Helvie, Janet Bailey, Lubomir M Hadjiiski
Department of Radiology, The University of Michigan, Ann Arbor MI

Purpose: We have previously developed an automated computer classifier for characterization of breast masses in 3D ultrasound volumes. Our purpose in this study was to investigate if this classifier would improve radiologists' accuracy in differentiation of malignant and benign breast masses.

Methods and Materials: The 3D volumes were recorded digitally as cine-clips when the transducer was translated across the lesion while conventional 2D images were acquired at each transducer location. Compared to 2D images, 3D ultrasound may provide additional information both to the radiologist and the computer. To take advantage of the additional information, the computer algorithm was designed to automatically delineate the mass boundaries in 3D, and to automatically extract features based on the segmented mass shapes and margins. The features were merged into a malignancy score using a computer classifier. The data set for the ROC study consisted of 102 ultrasound volumes from different patients containing biopsy-proven masses (44 benign and 58 malignant). None of the masses were simple cysts. The location of the biopsied lesion was identified by an experienced radiologist on all images. Five other MQSA radiologists participated as observers in the ROC study. Each radiologist read the cases first without CAD, immediately followed by reading with CAD. When reading without CAD, a radiologist analyzed the mass in the 3D volume using a graphical interface, and provided an estimate of the likelihood of malignancy. Subsequently, the radiologist was presented the computer malignancy score and had an option to revise his/her malignancy estimate. The reading order of the ultrasound volumes was randomized for each observer. The classification accuracy was quantified by using the area under ROC curve, Az.

Results: The computer classifier achieved an Az value of 0.92. The radiologists had an average Az of 0.84 (range: 0.82 to 0.86) without CAD. The accuracy of every radiologist was improved when they read with CAD (range: 0.04 to 0.11), and the average Az of the four radiologists improved to 0.90 (range: 0.87 to 0.93). The improvement was statistically significant ($p=0.006$).

Conclusion: A well-trained computer algorithm may improve radiologists' accuracy in characterizing breast masses as malignant and benign on ultrasound images.

APPENDIX 3

DEVELOPMENT OF A COMPUTER CLASSIFIER FOR COMPUTER-AIDED CHARACTERIZATION OF BREAST MASSES IN 3D ULTRASOUND VOLUMES

Berkman Sahiner, Heang-Ping Chan, Chintana Paramagul, Alexis V Nees, Caroline E Blane, Aditya Ramachandran
Department of Radiology, The University of Michigan, Ann Arbor MI

3D US volumes were recorded by translating the transducer across the lesion while conventional 2D images were acquired in the image planes. The basic steps in computer classifier design include lesion segmentation, feature extraction, and feature classification. We have developed 2D and 3D segmentation methods to delineate the mass boundaries in 3D US images. We extracted features that mimic those used by radiologists for malignant-benign classification, such as the width-to-height ratio and shadowing, as well as features that are less intuitive, such as the texture within the mass margins. Features were merged into a malignancy score using a linear classifier. 3D volumes containing biopsy-proven solid breast masses were collected from 102 patients (44 benign and 58 malignant). The area Az under the receiver operating characteristic curve for testing the computer classifier was 0.92. In comparison, the average Az for four MQSA radiologists who read the same cases without computer aid was 0.86. Our results indicate that an accurate computer classifier can be designed for differentiation of malignant and benign solid breast masses in 3D US volumes.

APPENDIX 4

Three-dimensional active contour model for characterization of solid breast masses on three-dimensional ultrasound images

Berkman Sahiner[*a], Aditya Ramachandran[a], Heang-Ping Chan[a], Marilyn Roubidoux[a], Lubomir M. Hadjiiski[a], Mark A. Helvie[a], Nicholas Petrick[b], Chuan Zhou[a]
[a]Department of Radiology, University of Michigan, Ann Arbor
[b]Center for Devices and Radiological Health, U.S. Food and Drug Administration, Rockville, MD

ABSTRACT

The accuracy of discrimination between malignant and benign solid breast masses on ultrasound images may be improved by using computer-aided diagnosis and 3-D information. The purpose of this study was to develop automated 3-D segmentation and classification methods for 3-D ultrasound images, and to compare the classification accuracy based on 2-D and 3-D segmentation techniques. The 3-D volumes were recorded by translating the transducer across the lesion in the z-direction while conventional 2-D images were acquired in the x-y plane. 2-D and 3-D segmentation methods based on active contour models were developed to delineate the mass boundaries. Features were automatically extracted based on the segmented mass shapes, and were merged into a malignancy score using a linear classifier. 3-D volumes containing biopsy-proven solid breast masses were collected from 102 patients (44 benign and 58 malignant). A leave-one-out method was used for feature selection and classifier design. The area A_z under the test receiver operating characteristic curves for the classifiers using the 3-D and 2-D active contour boundaries were 0.88 and 0.84, respectively. More than 45% of the benign masses could be correctly identified using the 3-D features without missing a malignancy. Our results indicate that an accurate computer classifier can be designed for differentiation of malignant and benign solid breast masses on 3-D sonograms.

Keywords: Computer-aided diagnosis, 3-D ultrasound, breast masses, segmentation, lesion classification

1. INTRODUCTION

The positive predictive value of biopsy recommendations for mammographically suspicious nonpalpable breast masses is between 20-30%.[1,2] Ultrasound imaging is a safe and inexpensive modality for further evaluation of masses detected mammographically or by palpation. Sonography has been shown to provide excellent accuracy for characterization of masses as cystic or solid.[3] In the 1980's and early 1990's, the overlap between malignant and benign features for non-cystic breast masses has prompted many investigators to recommend sonographic analysis only to determine whether the mass is cystic or solid.[3-5] However, a number of recent studies with state-of-the-art sonographic scanners have demonstrated that the characterization of non-cystic masses can be significantly improved by using ultrasound as an adjunct to mammography.[6-8]

Computer-aided diagnosis (CAD) methods can potentially improve the accuracy of breast cancer diagnosis by providing an unbiased and accurate second opinion to radiologists. It has been shown that the use of CAD can significantly improve radiologists' characterization of breast masses[9] and microcalcifications[10] as malignant or benign on mammograms. The increasing use of sonography to further evaluate and characterize both cystic and solid breast masses has prompted a number of researchers to investigate the application of CAD to breast ultrasound images to improve the characterization accuracy.[11-15]

3-D ultrasonography is rapidly gaining popularity as it moves out of the research environment and into the clinical setting.[16] Current technology allows radiologists to obtain 3-D or volumetric sonograms during clinical examination. We believe that computerized analysis of 3-D ultrasound images is important for two reasons. First, 3-D or volumetric ultrasound data may be more time-consuming for a radiologist to interpret, thus making CAD more desirable. Second, 3-D or volumetric ultrasound provides more data and better statistics, which should improve statistical image analysis.

* berki@umich.edu, phone 734-647-7429, CGC B2102, 1500 E. Medical Center Dr., Ann Arbor, MI 48109-0904

At our institution, a 3-D ultrasound image acquisition system was developed and was applied to imaging studies of several organs including the breast.[17,18] In our previous work using a limited data set, we investigated the use of texture features extracted from 3-D ultrasound images for characterization of breast masses as malignant or benign,[11] and the segmentation of these lesions using a 2-D active contour model.[13] In this study, we investigated the use of a 3-D active contour model for improved segmentation, and compared the computer characterization results based on segmentation using the 2-D and 3-D models on a larger data set.

2. METHODS

2.1 Image Acquisition

A 3-D ultrasound image acquisition system was previously developed and tested at our institution. The 3-D system consists of a commercially available ultrasound scanner (GE Logiq 700 with a M12 linear array transducer), a mechanical transducer guiding system, and a computer workstation. The 3-D data were acquired by translating the transducer in the cross-plane, or the z-direction, while acquiring conventional 2-D images in the x-y plane. The 2-D images were obtained at approximately equal incremental translations, which were measured and recorded using a translation sensor. The number of 2-D slices that were obtained was typically around 90, and varied depending on the lesion size. The maximum distance between two 2-D slices was 0.5 mm. The linear array transducer was operated at 11 MHz.

Before 3-D image acquisition, scout images were acquired to localize the lesion. During 3-D image acquisition, the technologist manually translated the transducer, while the image acquisition system recorded B-mode images into a buffer in the ultrasound scanner. After data acquisition, the images and the position data were transferred digitally to a workstation. The depth of the scans was kept constant at 3 or 4 cm for most of the patients. The technologist was free to set the focal distance and the overall gain adjustment to obtain the best possible image.

2.2 Data Set

Ultrasound images from 102 patients were included in this study. Patients were selected after identification of a mass based on mammography and/or clinical examination. Masses that proved to be simple cysts based on ultrasonography were excluded. All patients underwent biopsy as part of their clinical management after ultrasound imaging. Fifty-eight of the masses were malignant and 44 were benign.

The likelihood of malignancy for each mass, based on the 3-D ultrasound image, was rated by an expert radiologist on a scale of 1 to 100. In order to identify the biopsied mass, the radiologist was free to use the clinical ultrasound images and mammograms of the patients, as well as the radiology and pathology reports. However, the radiologist's malignancy rating was based on the appearance of the mass on the 3-D ultrasound image alone. This allowed us to compare the distribution of the malignancy ratings by the computer and by the radiologist based on the same information. The distribution of the ratings for the malignant and benign masses is shown in Figure 1.

In addition to providing a likelihood of malignancy rating, the radiologist was asked to fit a three-dimensional ellipsoid to the identified lesion. The best fit was obtained by scaling, rotating, and translating an ellipsoid superimposed on the 3-D data set using a dynamic object manipulation tool developed for this purpose. Figures 2a and 2b show five original consecutive slices and the radiologist-fitted ellipsoid for a mass that was seen in 14 slices in the 3-D volume.

2.3 2-D Active Contour Model

An active contour is a deformable continuous curve, whose shape is controlled by internal forces (the model, or *a-priori* knowledge about the object to be segmented) and external forces (the image).[19] The internal forces impose a smoothness constraint on the contour, and the external forces push the contour towards image edges. To solve a segmentation problem, an initial boundary is iteratively deformed so that the energy due to internal and external forces is minimized along the contour.

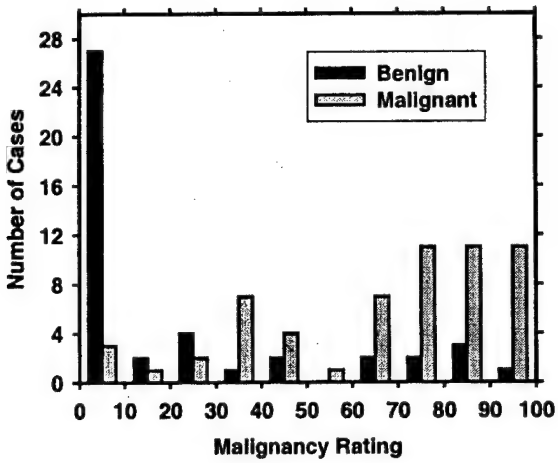

Figure 1: The distribution of the malignancy ranking of the masses in our data set, by an experienced radiologist. 1: very likely benign, 100: very likely malignant.

The internal energy components in our current active contour model are the continuity and curvature of the contour. The external energy components are the negative of the smoothed image gradient magnitude, and a balloon force that exerts pressure at a normal direction to the contour. The contour is represented by an N-point polygon whose vertices are $v(i) = (x(i), y(i)), i = 1, \ldots N$. The energy $E(i)$ for a vertex i is defined as a weighted average of the energy terms due to curvature, continuity, gradient and the balloon force at that vertex

$$E(i) = w_{curv} E_{curv}(i) + w_{cont} E_{cont}(i) + w_{grad} E_{grad}(i) + w_{bal} E_{bal}(i) \tag{1}$$

where w represents the weight, and the individual energy terms are defined in the next paragraph. The total energy to be minimized for the contour is defined as the sum of the energies for all vertices, i.e., $E = \sum_{i=1}^{N} E(i)$.

The curvature energy term is approximated by the second derivative of the contour, $E_{curv}(i) = \left|\mathbf{v}(i-1) - 2\mathbf{v}(i) + \mathbf{v}(i+1)\right|$. This term is large when the angle at vertex i is small. By discouraging small angles at vertices, this term attempts to smooth the contour. The continuity term is represented by the deviation of the length of the line segment under consideration from the average line segment length $\bar{d}$. Therefore this term helps the vertices maintain regular spacing along the contour. The image gradient magnitude is obtained by first smoothing the image with a low-pass filter, then computing a partial derivative vector whose components are the derivatives in the horizontal and vertical directions, and finally computing magnitude of the partial derivative vector. Since the gradient energy is defined as the negative of the gradient magnitude, minimizing this term attracts the contour to image edges. The balloon energy term is required to prevent the contour from collapsing onto itself, which is a well-known phenomenon in active contour models.[20]

To minimize the contour energy, we used a greedy algorithm that was first proposed by Williams and Shah.[21] In this algorithm, the contour was iteratively optimized starting with an initial contour. At each iteration, a neighborhood of each vertex was examined, and the vertex was moved to the location that minimizes the contour energy. The algorithm was terminated when there was no movement of the vertices, or when the ratio of moved vertices to the total number of vertices was less than an input threshold. The algorithm was initialized with the radiologist-fitted ellipsoid discussed in Section 2.2. Figure 2c shows the final contours obtained with the 2-D model.

2.4 3-D Active Contour Model

Since the 2-D active contour model acts on each slice of the 3-D volume independently, the segmented regions on neighboring slices can be quite different, as seen in Figure 2c. However, it is known that the shape of the mass is

unlikely to change dramatically from one slice to the next, because the spacing between these slices is relatively small (at most 0.5 mm). This discontinuity in the mass shape across adjacent slices is a result of the fact that the 2-D active contour does not make full use of the 3-D data. Our 3-D active contour model is aimed at using the shape information across the 3-D slices to improve upon the 2-D active contour model.

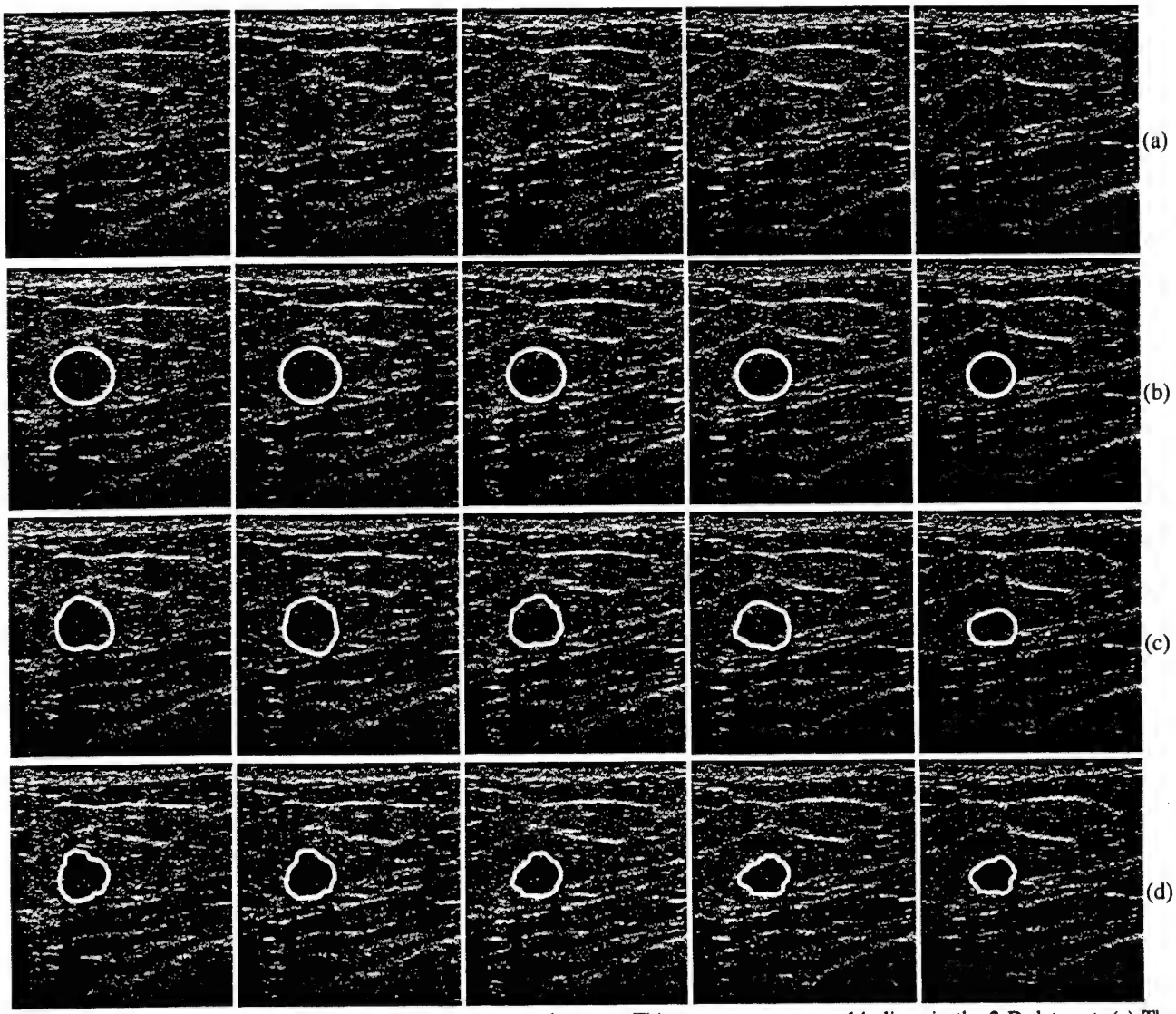

Figure 2: Five consecutive slices of a 3-D mass from our data set. This mass was seen on 14 slices in the 3-D data set. (a) The original image (b) Radiologist-fitted ellipsoid, (c) The result of the 2-D active contour segmentation, (d) The result of the 3-D active contour segmentation.

Our 3-D active contour model is defined by including in the curvature energy term an additional component related to the smoothness of the mass in the z-direction. Let *v(i,j)* denote the i^{th} vertex in image slice j. The curvature energy in our 3-D active contour model is defined as

$$E_{curv}(i,j) = w_{cip}\left|\mathbf{v}(i-1,j) - 2\mathbf{v}(i,j) + \mathbf{v}(i+1,j)\right| + w_{cop}\left|\mathbf{v}(i,j-1) - 2\mathbf{v}(i,j) + \mathbf{v}(i,j+1)\right|, \quad (2)$$

where the subscripts *cip* and *cop* stand for "curvature in-plane", and "curvature out-of-plane", respectively. The second component in the curvature energy term forces the contour to be smooth in the z direction. Figures 2c and 2d compare the 2-D and 3-D active contour segmentation results. Figure 3 shows the entire 3-D object segmented on 14 slices.

2.5 Feature extraction and classification

The features used in this study were extracted from spatial gray level dependence (SGLD) matrices derived from 2-D slices of the 3-D data set. Features extracted from SGLD matrices of ultrasound images have been shown to be useful in classification of malignant and benign breast masses in previous studies.[22]. Six texture feature measures that are invariant under linear, invertible gray scale transformations were extracted. These features were information measures of correlation 1 and 2 (IMC1 and IMC2), energy (ENE), entropy (ENT), sum entropy (SME) and difference entropy (DFE). Although many gray scale transformations will not be invertible due to pixel saturation or roundoff, these features will reduce the effect of gray-level adjustments on the classification accuracy.

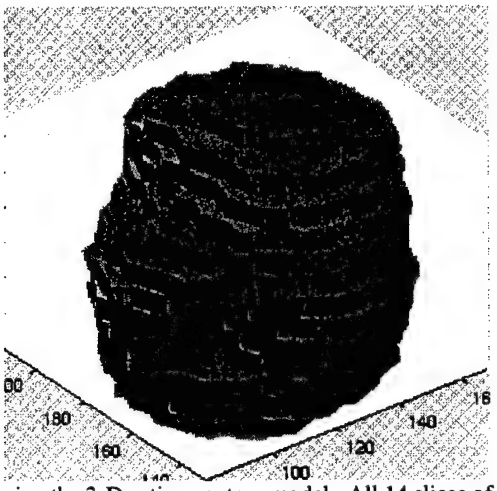

Figure 3: The segmented mass object using the 3-D active contour model. All 14 slices of the mass shown in Figure 2 are included.

It is known that the margin characteristics of a mass are very important for its characterization. For this reason, the texture features were extracted from two disk-shaped regions containing the boundary of each mass, as well as mass and normal tissue adjacent to the boundary of the mass. These regions followed the contour determined by the active contour model, as shown in Figure 4. The areas for the upper and lower disk-shaped regions were chosen to be equal, and their sum was equal to the area of the segmented mass. The pixel-pair distances used for SGLD matrix computation were d=2, 4, and 6. Two pixel-pair angles, $\theta=0^\circ$ and $\theta=90^\circ$ were evaluated for both regions. After a given texture feature was extracted from the different 2-D slices of a given mass, a 3-D feature was obtained by averaging the feature values for the 2-D slices. The final feature space for the classifier therefore contained 6 feature measures from 2 regions at 3 pixel-pair distances and 2 pixel-pair angles, which resulted in a 72-dimensional feature space.

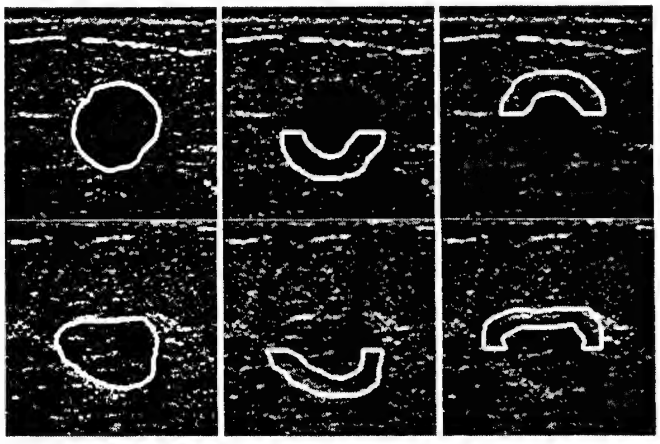

Figure 4: The active contour segmentation, and the upper and lower disk-shaped regions for two masses in our data set.

Stepwise feature selection and linear discriminant analysis were used for classifier design. A leave-one-case-out methodology was used to train and test the classifier with N=102 cases. The training included feature selection and the computation of classifier coefficients for the selected features using N-1 cases. The test scores were analyzed using receiver operating characteristic (ROC) methodology. The classification accuracies using features obtained from the 2-D and 3-D active contour segmentation were compared.

3. RESULTS

The stepwise feature selection method selected an average of 5 and 7 features for 2-D and 3-D active contour segmentation, respectively. Figures 5a and 5b show the distribution of the classifier test scores for the malignant and benign cases using the two segmentation methods. By choosing an appropriate threshold on the test scores of the classifier designed using the 2-D segmentation model, 11% of the benign masses could be correctly identified without missing a malignant mass. This number was improved to 45% with the use of the 3-D segmentation model. Figure 6 shows the ROC curves for the classifiers designed using features extracted from contours provided by the two segmentation methods. The area A_z under the ROC curve was 0.84 and 0.88 for 2-D active contour segmentation and 3-D active contour segmentation, respectively. When the radiologist's malignancy rating was used as the decision variable in ROC analysis, an A_z value of 0.84 was obtained.

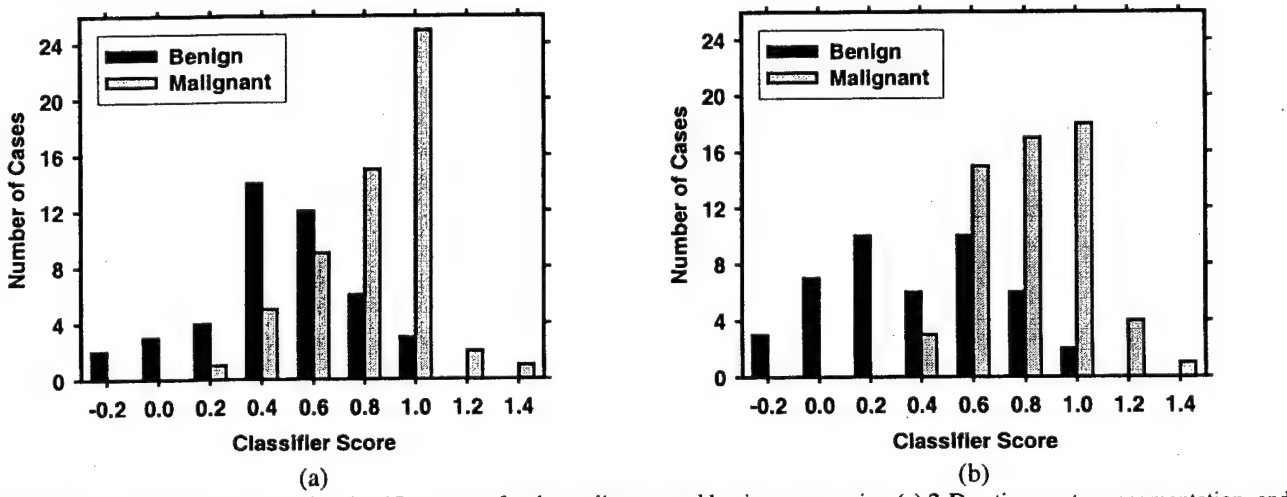

Figure 5: The distribution of the classifier scores for the malignant and benign cases using (a) 2-D active contour segmentation, and (b) 3-D active contour segmentation.

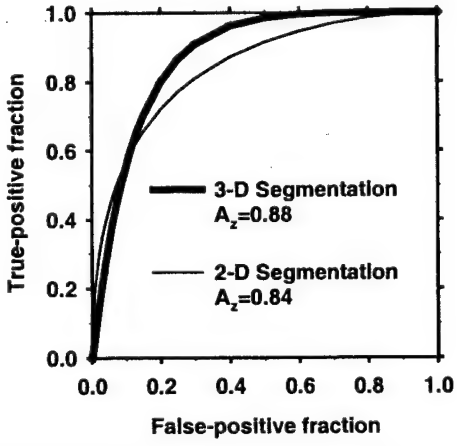

Figure 6: The ROC curves for the classifiers designed using features extracted from the 2-D active contour segmentation and 3-D active contour segmentation.

4. DISCUSSION AND CONCLUSION

Our results indicate that solid breast masses can be accurately classified as malignant or benign by computerized analysis of 3-D ultrasound images. The computer classifier using the 3-D segmentation method achieved an A_z value of 0.88 in the task of characterizing solid breast masses. An experienced radiologist's A_z value for the same task and using the same information was 0.84. Although the difference between the characterization accuracy of the computer and that of the radiologist did not achieve statistical significance, our preliminary analysis indicates that the computer's performance may be similar to that of an experienced radiologist. We plan to conduct further observer studies with more radiologists to confirm this analysis.

We have observed that the 3-D active contour model can be more accurate than the 2-D active contour model in the task of segmenting masses on 3-D sonograms. An example comparing 3-D segmentation with 2-D segmentation for the same mass is provided in Figure 2. In Figure 2c, the 2-D method delineates the upper boundary of the mass adequately in the first and last slices shown (leftmost and rightmost columns, respectively). However, in the second, third, and fourth columns, the upper boundary overestimates the size of the mass. Since the active contour is optimized independently for each slice, the correct information from the first and last slices in this figure cannot be used to improve the segmentation for the middle slices. The 3-D model, which uses the information from multiple slices, succeeds in delineating the mass boundaries more accurately, as shown in Figure 2d. The characterization results also confirm that 3-D segmentation may be more accurate, although the difference between the characterization accuracy with features extracted from the 2-D active contour segmentation (A_z=0.84) and that with the 3-D active contour segmentation (A_z=0.88) did not achieve statistical significance.

We plan to perform observer performance studies to investigate the effect of our computer classifier on the radiologists' accuracy in characterizing malignant and benign solid masses on 3-D sonograms. The potential of the developed classifier cannot be fully realized until the sonographic analysis is used in conjunction with mammographic analysis. The combination of the computerized characterization results from these two modalities is under investigation.

ACKNOWLEDGMENTS

This work was supported in part by a U.S. Army Medical Research Materiel Command grant DAMD17-01-1-0328 and by USPHS grant CA91713. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred.

REFERENCES

1. F. M. Hall, J. M. Storella, D. Z. Silverstone, and G. Wyshak, "Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography," *Radiology* **167**, 353-358, 1988.

2. M. L. Brown, F. Houn, E. A. Sickles, and L. G. Kessler, "Screening mammography in community practice: positive predictive value of abnormal findings and yield of follow-up diagnostic procedures," *Amer. J. Roentgenology* **165**, 1373-1377, 1995.

3. V. P. Jackson, "The role of US in breast imaging," *Radiology* **177**, 305-11, 1990.

4. C. Cole-Beuglet, R. Z. Soriano, A. B. Kurtz, and B. B. Goldberg, "Ultrasound analysis of 104 primary breast carcinomas classified according to histopathologic type," *Radiology* **147**, 191-196, 1983.

5. L. A. Venta, C. M. Dudiak, C. G. Salomon, and M. E. Flisak, "Sonographic evaluation of the breast," *Radiographics* **14**, 29-50, 1994.

6. A. T. Stavros, D. Thickman, C. L. Rapp, M. A. Dennis, S. H. Parker, and G. A. Sisney, "Solid breast nodules: Use of sonography to distinguish between malignant and benign lesions," *Radiology* **196**, 123-134, 1995.

7. P. Skaane and K. Engedal, "Analysis of sonographic features in differentiation of fibroadenoma and invasive ductal carcinoma," *AJR* **170**, 109-114, 1998.

8. K. J. W. Taylor, C. Merritt, C. Piccoli, R. Schmidt, G. Rouse, B. Fornage, E. Rubin, D. Georgian-Smith, F. Winsberg, B. Goldberg, and E. Mendelson, "Ultrasound as a complement to mammography and breast examination to characterize breast masses," *Ultrasound Med. Biol.* **28**, 19-26, 2002.

9. H.-P. Chan, B. Sahiner, M. A. Helvie, N. Petrick, M. A. Roubidoux, T. E. Wilson, D. D. Adler, C. Paramagul, J. S. Newman, and S. S. Gopal, "Improvement of radiologists' characterization of mammographic masses by computer-aided diagnosis: an ROC study," *Radiology* **212**, 817-827, 1999.

10. Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger, and K. Doi, "Improving breast cancer diagnosis with computer-aided diagnosis," *Acad. Radiol.* **6**, 22-33, 1999.

11. B. Sahiner, G. L. LeCarpentier, H.-P. Chan, M. A. Roubidoux, N. Petrick, M. M. Goodsitt, S. S. Gopal, and P. L. Carson, "Computerized characterization of breast masses using three-dimensional ultrasound images," *Proc. SPIE* **3338**, 301-312, 1998.

12. D. R. Chen, R. F. Chang, and Y. L. Huang, "Computer-aided diagnosis applied to US of solid breast nodules by using neural networks," *Radiology* **213**, 407-412, 1999.

13. B. Sahiner, H.-P. Chan, G. L. LeCarpentier, N. Petrick, M. A. Roubidoux, and P. L. Carson, "Computerized characterization of solid breast masses using three-dimensional ultrasound images," *Radiology* **213(P)**, 229, 1999.

14. K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny, "Computerized diagnosis of breast lesions on ultrasound," *Med. Phys.* **29**, 157-164, 2002.

15. C. M. Chen, Y. H. Chou, K. C. Han, G. S. Hung, C. M. Tiu, H. J. Chiou, and S. Y. Chiou, "Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks," *Radiology* **226**, 504-514, 2003.

16. D. B. Downey, A. Fenster, and J. C. Williams, "Clinical utility of three-dimensional US," *Radiographics* **20**, 559-571, 2000.

17. P. L. Carson, A. P. Moskalik, A. Govil, M. A. Roubidoux, J. B. Fowlkes, D. Normolle, D. D. Adler, J. M. Rubin, and M. A. Helvie, "The 3D and 2D color flow display of breast masses," *Ultrasound Medical Biol.* **23**, 837-849, 1997.

18. P. T. Bhatti, G. L. LeCarpentier, M. A. Roubidoux, J. B. Fowlkes, M. A. Helvie, and P. L. Carson, "Discrimination of sonographically detected breast masses using frequency shift color Doppler imaging in combination with age and gray scale criteria," *Journal of Ultrasound in Medicine* **20**, 343-350, 2001.

19. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Int. J. Comput. Vision* **1**, 321-331, 1987.

20. L. D. Cohen, "On active contour models and balloons," *CVGIP: Img. Underst.* **53**, 211-218, 1991.

21. D. J. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation," *CVGIP: Img. Underst.* **55**, 14-26, 1992.

22. B. S. Garra, B. H. Krasner, S. C. Horri, S. Ascher, S. K. Mun, and R. K. Zeman, "Improving the distinction between benign and malignant breast lesions: The value of sonographic texture analysis," *Ultrasonic Imaging* **15**, 267-285, 1993.

APPENDIX 5

Computerized Characterization of Breast Masses on 3D Ultrasound Volumes

Berkman Sahiner
Heang-Ping Chan
Marilyn M. Roubidoux
Mark A. Helvie
Lubomir M. Hadjiiski
Aditya Ramachandran
Chintana Paramagul
Gerald L. LeCarpentier
Alexis Nees
Caroline Blane

Department of Radiology
University of Michigan, Ann Arbor

Short title: Characterization of Breast Masses on 3D Ultrasound

Correspondence:
Berkman Sahiner, Ph.D.
Department of Radiology
University of Michigan
1500 E. Medical Center Drive
CGC B2102
Ann Arbor, MI 48109-0904
Telephone: (734) 647-7429
Fax: (734) 615-5513
e-mail: berki@umich.edu

ABSTRACT

We are developing computer vision techniques for characterization of breast masses as malignant or benign on radiologic examinations. In this study, we investigated computerized characterization of breast masses on 3D ultrasound (US) volumetric images. We developed 2D and 3D active contour models for automated segmentation of the mass volumes. The effect of the initialization method of the active contour on the robustness of the iterative segmentation method was studied by varying the contour used for its initialization. For a given segmentation, texture and morphological features were automatically extracted from the segmented masses and their margins. Stepwise discriminant analysis with the leave-one-out method was used to select effective features for the classification task and to combine these features into a malignancy score. The classification accuracy was evaluated using the area A_z under the receiver operating characteristic (ROC) curve, as well as the partial area index $A_z^{(0.9)}$, defined as the relative area under the ROC curve above a sensitivity threshold of 0.9. For the purpose of comparison with the computer classifier, four experienced breast radiologists provided malignancy ratings for the 3D US masses. Our data set consisted of 3D US volumes of 102 biopsied masses (44 benign, 58 malignant). The classifiers based on 2D and 3D segmentation methods achieved test A_z values of 0.88±0.03 and 0.92±0.03, respectively. The difference in the A_z values of the two computer classifiers did not achieve statistical significance. The A_z values of the four radiologists ranged between 0.84 and 0.92. The difference between the computer's A_z value and that of any of the four radiologists did not achieve statistical significance. However, the computer's $A_z^{(0.9)}$ value was significantly higher than that of three of the four radiologists. Our results indicate that an automated and effective computer classifier can be designed for differentiating malignant and benign breast masses on 3D US volumes. The accuracy of the classifier designed in this study was similar to that of experienced breast radiologists.

Keywords: computer-aided diagnosis, 3D ultrasound, breast mass characterization, segmentation

I. INTRODUCTION

The importance of early breast cancer detection requires a vigorous approach to characterization of breast lesions. At present, the positive biopsy rate for nonpalpable breast lesions as well as for nonpalpable breast masses is between 15-30%.[1-4] This means that 70-85% of breast biopsies are performed for benign lesions. In order to reduce patient anxiety and morbidity, as well as to decrease health care costs, it is desirable to reduce the number of benign biopsies without missing malignancies. Computer-aided diagnosis (CAD) can provide a consistent and reproducible second opinion to the radiologists, and has a potential to assist them in reducing benign biopsies. Recent studies on the computerized classification of breast masses based on mammographic image features suggest that radiologists' performance may be significantly improved if they are aided by a well-trained CAD system.[5-7] Breast ultrasound (US) is an important imaging modality for characterization of breast masses as malignant and benign. An objective and reproducible second opinion from a computer classifier for classification of breast masses based on US image features may be an important addition to CAD tools being developed for mammographic image analysis.

Breast US is widely accepted as a highly accurate modality for the differentiation of cystic and non-cystic masses. As a result of technological improvements and more sophisticated utilization by radiologists, US has been gaining popularity for characterization of non-cystic, or solid, breast masses. By combining several ultrasonic characteristics, Stavros et al.[8] achieved a specificity of 98.4% and a sensitivity of 68.7% on a data set of 750 solid breast masses. Using strict criteria for a benign diagnosis, Skaane et al.[9] achieved a positive predictive value of 66% and a negative predictive value of 98% for differentiation of fibroadenoma and invasive ductal carcinoma on sonograms. Recently, Taylor et al. investigated whether the complementary use of US imaging could decrease the biopsy of benign, non-cystic masses. On a data set of 761 biopsied masses, they found that the addition of US evaluation to mammography alone could increase the specificity from 51.4% to 63.8% while slightly increasing the sensitivity from 97.1% to 97.9%.[10] Our study aims at developing techniques for computerized characterization of solid breast masses, which may eventually improve radiologists' accuracy in this difficult and important task.

A number of researchers have recently investigated the application of CAD to breast US images.[11-14] D.R. Chen et al.[12] extracted autocorrelation features from rectangular regions of interest (ROIs) containing solid breast masses. Using a neural network classifier, they obtained an area A_z under the receiver operating characteristic (ROC) curve of 0.956 for classification of a data set of 140 biopsy-proven masses as malignant or benign. Horsch et al.[13] developed an automated segmentation method for delineating the mass boundaries, and compared its characterization accuracy on different subsets with that obtained from manual segmentation. Using manual and automated segmentation methods, they obtained A_z values of 0.91 and 0.87, respectively, in the task of differentiating all malignant and benign lesions in their data set, and 0.88 and 0.82, respectively, in the task of differentiating the subset of malignant and benign solid lesions. C.M. Chen et al.[14] used morphological features extracted from manually segmented mass boundaries for classification. Using a neural network classifier, they obtained an A_z of 0.959 for classification of a data set of 271 biopsy-proven masses as malignant or benign.

3D US is rapidly gaining popularity as it moves out of the research environment and into the clinical setting.[15] Computerized analysis of 3D US images may be useful for two reasons. First, 3D or volumetric US data may be more time-consuming for a radiologist to interpret, thus making CAD more desirable. Second, 3D or volumetric US provides more data and better statistics, which should improve statistical image analysis.

In clinical practice, breast US may be performed in different ways. In many breast imaging clinics, the US examination is performed by a US technologist. Once the technologist locates the mass, and determines the appropriate settings for optimal image quality, representative static US images of the mass are printed on hardcopy film. The radiologist only reads the images chosen by the technologist. A second possibility is that the US scan is videotaped by the technologist and the radiologist reads the examination on video display. In a third method, a radiologist will perform the US examination interactively and optimize the image quality by changing the probe angle, direction, and US machine settings. Since the US image quality is operator dependent, the way in which the examination is performed may have an impact on the diagnostic accuracy. At our institution, the third method is employed. As described in the Methods Section, the data acquisition system in this study did not permit interactive modification during 3-D image acquisition. As a result, the data that was used by the computer and the radiologists for mass characterization in this study may not be as informative as the

data that the radiologists could have obtained by examining the patient interactively. However, since the mass is entirely imaged in the 3D data set, our data should be at least comparable to that obtained by using the first method described above.

In this study, we investigated the computerized characterization of non-cystic breast masses as malignant and benign in 3D US images. We developed a 3D segmentation method to delineate the masses. Morphological and texture features were extracted from the mass and its margins for classification. A linear classifier was used to merge the features into a malignancy score. The classification accuracy was evaluated by ROC methodology. The ROC curves of the computer and four experienced breast radiologists were compared. To our knowledge, this is the first study on 3D US images that investigates a computer segmentation method followed by a computer classifier for breast cancer characterization.

II. METHODS

A. Data Set

Institutional review board approval was obtained prior to the commencement of this investigation. The images used in this study were acquired between 1998 and 2002. Our study group was 102 women (average age: 51 years) who had a solid mass deemed suspicious or highly suggestive of malignancy. All patients underwent biopsy or fine needle aspiration. Fifty-eight masses were malignant and 44 were benign. Forty-five of the malignancies were invasive ductal carcinoma, five were invasive lobular carcinoma, one was medullary carcinoma, three were ductal carcinoma in-situ, and four were other invasive carcinoma. Of the benign masses, the majority were fibroadenoma (N=17) and fibrocystic disease (N=11). The mean equivalent lesion diameter was 1.28 cm (standard deviation = 0.78cm).

The 3D US data were acquired using an experimental system that was previously developed and tested at our institution.[16,17] The 3D system consisted of a commercially available US scanner (GE Logiq 700 with an M12 linear array transducer), a mechanical transducer guiding system, and a computer workstation. The linear array transducer was operated at 11 MHz. The technologist was free to set the focal distance and the overall gain adjustment to obtain the best possible image. Before 3D

image acquisition, the technologist used clinical US and mammogram images to identify the suspicious mass. . During 3D image acquisition, the technologist manually translated the transducer linearly in the cross-plane, or the z-direction, while the image acquisition system recorded 2D B-mode images in the image scan plane (x-y plane). The 2D images were obtained at 0.5 mm incremental translations, which were measured and recorded using a translation sensor. The number of 2D slices was typically around 90, and varied depending on the lesion size. The maximum distance between two 2D slices was 0.5 mm. The scanned breast region measured typically 4.5 cm long by 4.0 cm wide by 4.0 cm deep. The typical pixel size in a slice was approximately 0.11 mm.

The B-mode images were recorded into a buffer in the US scanner. After data acquisition, the images and the position data were transferred digitally to a workstation, where individual planes were cropped and stacked to form a 3D volume. The biopsied mass in each volume was identified by an MQSA (Mammography Quality Standards Act) qualified radiologist (RAD1) using clinical US and mammographic images to confirm that the 3D images contained the clinically suspicious mass. The likelihood of malignancy for each mass, based on the 3D US image alone, was rated by the same radiologist on a scale of 1 to 100, where a higher number corresponded to a higher likelihood of malignancy. The distribution of the ratings for the malignant and benign masses is shown in Fig. 1. The radiologist was also asked to fit a 3D ellipsoid to the mass. The 3D ellipsoid was used to initialize the computerized mass segmentation described in the next section. The best fit was obtained by scaling, rotating, and translating an ellipsoid superimposed on the 3D data set using a dynamic object manipulation tool developed for this purpose.

B. Mass Segmentation

We investigated the use of 2D and 3D active contour models for segmentation of mass boundaries.[18] An active contour model is a high-level segmentation method that uses energy terms derived from the image gray-level information as well as the *a-priori* knowledge about the object to be segmented for accurate segmentation. The segmentation problem is defined as an energy minimization problem. In order for the model to lock onto the contours in the image, the image-based energy terms, also referred to as the external energy terms, are usually defined in terms of the image gray levels and the image gradient magnitude. *A-priori* knowledge of object shape is used to define internal energy terms related to features such as the continuity and the smoothness of the contour to constrain the segmentation problem.

These terms can compensate for noise or apparent gaps in the image gradients, which often mislead segmentation methods that do not use *a-priori* information.

In a 2D segmentation problem, the contour of the object can be represented by N vertices, $v(i) = ((x(i), y(i)), i = 1,..., N$. In the discrete formulation of the active contour model, the total energy to be minimized is defined as $E = \sum_{i=1}^{N} E(i)$, where $E(i)$ is the energy at vertex i, defined as the sum of the internal and external energy terms $E_k(i)$, given by $E(i) = \sum_k w_k E_k(i)$, and w_k is the weight of the i^{th} energy term. In our 2D active contour model, we used four internal and external energy terms, namely, the gradient magnitude, continuity, smoothness, and balloon energy. The image gradient magnitude, $E_1(i)$, was obtained by first smoothing the image with a low-pass filter, then computing a partial derivative vector whose components are the derivatives of the filtered image in the horizontal and vertical directions, and finally computing the magnitude of the partial derivative vector. The weight of the gradient energy is defined to be a negative number; thus, minimizing $w_1 E_1$ attracts the contour to image edges. The continuity term is represented by the deviation of the length of the line segment $d(i)$ between vertices i and $i+1$ from the average line segment length $\bar{d}$, i.e., $E_2(i) = | d(i) - \bar{d} |$. Therefore minimizing this term helps the vertices maintain regular spacing along the contour. The curvature term, $E_3(i)$, is approximated by the second derivative of the contour: $E_3(i) = |\mathbf{v}(i-1) - 2\mathbf{v}(i) + \mathbf{v}(i+1)|$. As long as $| d(i) - \bar{d} |$ is small, this term will be large when the angle at vertex i is small. By discouraging small angles at vertices, this term attempts to smooth the contour. The balloon energy E_4 pushes the contour outward or pulls it inward, depending on whether w_4 is positive or negative, respectively, along a path normal to the contour. This energy term helps the active contour traverse spurious, isolated, or weak image edges, and counters its tendency to shrink. The resulting snake is reported to be more robust to the initial position and image noise.[19]

To solve the energy minimization problem, we have chosen the iterative method proposed by Williams and Shah.[20] The contour is first initialized by defining N vertices $v(i)$, $i=1,...,N$. At a given iteration, the method visits each vertex $v(i)$. Let $D(i)$ represent the set of pixels (x',y') in a

(2M+1)x(2M+1) neighborhood centered around $v(i)$. For each pixel in $\boldsymbol{D}(i)$, the sum $\sum_k w_k E_k$ is computed, and the vertex i is moved to the (x',y') location that minimizes this sum. After the minimization is performed locally at vertex $v(i)$, the algorithm moves to the vertex $v(i+1)$. The method converges when no vertex changes location at a given iteration. In practical implementation, iterations may be stopped when a large, predetermined percentage of vertices stop moving. The cross-section of the radiologist-defined ellipsoid with each image slice was used for initializing the contour.

When the 2D active contour model described above is applied to a 3D data set, segmentation is performed independently on each slice of the 3D volume. However, this kind of segmentation ignores the continuity of the object across slices. When the slice spacing is small compared to the rate of change of the object shape, it is known that the shape of the object is unlikely to change dramatically from one slice to the next. Our 3D active contour model is aimed at using the shape information across the 3D slices to improve upon the 2D active contour model. Our 3D active contour model is defined by including in the curvature energy term an additional component related to the smoothness of the mass in the z-direction. Let $v(i,j)$ denote the i^{th} vertex in image slice j. The curvature energy in our 3D active contour model is defined as

$$E_3(i,j) = w_{3\_in}\left|\mathbf{v}(i-1,j) - 2\mathbf{v}(i,j) + \mathbf{v}(i+1,j)\right| + w_{3\_out}\left|\mathbf{v}(i,j-1) - 2\mathbf{v}(i,j) + \mathbf{v}(i,j+1)\right|, \qquad (1)$$

where w_{3_in} and w_{3_out} stand for the weights for the in-plane and out-of-plane components of the curvature, respectively. The out-of-plane component in the curvature energy term forces the contour to be smooth in the z direction. Our implementation of the 3D active contour model starts by optimizing the contour on the first slice of the 3D data set (j=1). Since slice j=0 does not exist, we assume that $v(i,0) = v(i,1)$ for all i. After the contour is optimized for slice j=1, the optimization is performed for slice j=2, and so on, until the contour has been optimized for all slices. This constitutes one 3D iteration. The 3D model repeats 3D iterations until there is no movement of the vertices for the 3D contour, or when a predetermined percentage of vertices stop moving. Similar to our 2D active contour, the 3D active contour was initialized using the radiologist-defined ellipsoid.

C. Feature Extraction

We have evaluated a number of morphological and texture features for characterization of the masses as malignant or benign. Each of the features described below was extracted from every slice

where the mass was segmented using either the 2D or the 3D automated segmentation algorithm. The features extracted from different slices of the same mass were then combined to define the feature measures (such as mean or maximum) for that mass.

Extraction of morphological features

The taller-than-wide shape of a sonographic mass is a good indication of malignancy.[8] This characteristic was defined by the ratio of the widest cross section (W) of the automatically segmented lesion shape to the tallest cross section (T) in a slice (Fig. 2). Another feature that has been reported to be useful for differentiation of malignant and benign masses is posterior shadowing. In order to define a posterior shadowing feature (PSF), we first calculated the mean pixel value $\overline{R(i)}$ in overlapping vertical strips $R(i), i = 1,\ldots,n$ posterior to the mass, as shown in Fig. 2. The width W_R of a strip was equal to one fourth of the width of the mass ($W/4$), and the height of the strip was equal to the height of the mass (T). The left and right edges of strips $R(i)$ and $R(i+1)$ differed by one pixel. In other words, the strip $R(i+1)$ was obtained by moving the strip $R(i)$ to the right by one pixel, while, of course, the strip remained posterior to the mass and its height remained as T. In order to exclude the bilateral posterior shadowing artifacts that are sometimes associated with fibroadenomas, the strips were defined only posterior to the central $3W/4$ portion of the mass (Fig. 2). The minimum value of these averages, $\min\{\overline{R(i)}, i = 1,\ldots n\}$, was the darkest posterior strip. The PSF was defined as the normalized average gray-level difference between the interior of the segmented mass and the darkest posterior strip,

$$PSF = \frac{\overline{M} - \min\{\overline{R(i)}, i = 1,\ldots n\}}{\overline{M}}, \qquad (2)$$

where $\overline{M}$ denotes the mean gray level value inside the segmented mass.

Extraction of texture features

The features used in this study were extracted from spatial gray level dependence (SGLD) matrices derived from 2D slices of the 3D data set. The $(i,j)^{th}$ element of the co-occurrence matrix is the relative frequency with which two pixels, one with gray level i and the other with gray level j, separated by a pixel pair distance d in a direction θ occur in the image. Features extracted from SGLD matrices of US images have been shown to be useful in classification of malignant and benign breast masses on mammograms in previous studies.[21] In this study, six texture feature measures that are invariant under

linear, invertible gray scale transformations were extracted. These features were information measures of correlation 1 and 2 (IMC1 and IMC2), difference entropy (DFE), entropy (ENT), energy (ENE), and sum entropy (SME). The mathematical definitions of these features can be found in the literature.[22] Although many gray scale transformations may not be invertible due to pixel saturation or roundoff, these features are largely independent of the gray-level gain adjustments.

It is known that the margin characteristics of a mass are very important for its characterization, and previous studies have indicated that texture features extracted from the mass margins are effective for classification.[23] For this reason, the texture features in this study were extracted from two disk-shaped regions containing the boundary of each mass, as well as presumably mass and normal tissue adjacent to the boundary of the mass. These regions followed the contour determined by the active contour model, as shown in Fig. 3. The areas for the upper and lower disk-shaped regions were chosen to be equal, and their sum was equal to the area of the segmented mass. The pixel pair distances used for SGLD matrix computation were chosen to be d=2, 4, and 6. Two pixel pair angles, $\theta=0^\bullet$ and $\theta=90^\bullet$ were evaluated for each d in both regions. The number of SGLD matrices computed for a disk-shaped region was therefore 6, and the number of features extracted from an image containing the segmented mass was 72 (6 features, extracted from 6 SGLD matrices in the upper disk-shaped region and the lower disk-shaped region).

D. Classification

The features extracted from different slices of the same mass were combined to define the feature measures for that mass. For the width-to-height feature and the PSF, we computed the mean, variance, minimum and maximum of the extracted value from each slice containing the mass. Therefore eight morphological feature measures were defined for each mass. For texture features, we only computed the mean, hence 72 texture feature measures were defined for each mass.

Fisher's linear discriminant analysis (LDA)[24] was used for combining the features into a discriminant score. Since the number of available features in the feature space was relatively high compared with the number of available cases, stepwise feature selection[25] was used in order to reduce the number of the features and to obtain the best feature subset to design an effective classifier. For partitioning the data set into trainers and testers, we used the leave-one-case-out resampling method. Feature selection is

performed as part of the classifier design such that both the feature selection and the classifier coefficient estimation procedures were repeated 102 times, as each case were left out once as the test sample. The test discriminant scores were analyzed using ROC methodology.[26] The classification accuracy was evaluated using the area under the ROC curve, A_z, as well as the partial area index, $A_z^{(0.9)}$. $A_z^{(0.9)}$ is defined as the area under the ROC curve above a sensitivity threshold of 0.9 ($TPF_0 = 0.9$) normalized to the total area above TPF_0, which is equal to $(1\text{-}TPF_0)$.[27]

E. Malignancy Ranking by Radiologists

Although all the cases in our data set were suspicious enough to warrant biopsy or fine needle aspiration, the degree of difficulty of our cases can best be measured by investigating the accuracy of the radiologists in classifying the cases in our data set as malignant or benign. As described in Section II.B, one radiologist (RAD1) who was familiar with the clinically obtained images had initially provided a malignancy rating. To compare with the computer's accuracy, we are interested in measuring the accuracy of other radiologists, who would not be biased by memory or familiarity with the cases. For this purpose, we have developed an interactive graphical user interface with which the radiologists could navigate through 3D volumes, adjust the window and level of the displayed images, and enter a malignancy rating between 1 and 100 (higher rating indicating higher likelihood of malignancy) when they finish examining a case. Three additional radiologists (RAD2-RAD4) participated in the malignancy rating study. The radiologists RAD1-RAD4 were either fellowship-trained in breast imaging or had over 25 years of experience in breast imaging. All four radiologists were MQSA qualified and had experience ranging from 2 to 25 years (mean, 11.3 years of mammographic and US interpretation). The location of the center of mass, as determined by RAD1, was displayed on each slice, so that all the radiologists would rank the same mass if more than one mass existed in the volume. There was no time limitation for the radiologists to read a case. The case reading order was randomized for each radiologist. The malignancy rating was entered by means of a slide bar. Before participating in the study, the radiologists were trained on five cases that were not part of the test data set described in Section II.A. The malignancy rating study was intended to measure the difficulty of the data set, and was not intended to measure how the radiologists' interpretation would be affected by CAD. Therefore, the computer classification results were not displayed to the radiologists in this study.

III. RESULTS

We evaluated the accuracy of characterization based on both 2D and 3D active contour segmentation methods. Rows 1 to 4 of Fig. 4 show the original images, radiologist-defined ellipsoid, 2D active contour results, and 3D active contour results for five consecutive slices of a mass that was visible on a total of 10 slices. Fig. 5 shows a 3D rendering of the segmented object using the 2D and 3D active contour models. It is seen from Fig. 5 that the shape of the object segmented by the 3D active contour model is smoother in the z direction.

Table 1 shows the range (minimum and maximum) of the A_z values provided by each texture feature alone, extracted from the upper and lower disk-shaped regions determined by the 2D and 3D active contour models. The ranges in this table are for different pixel pair distances and directions used in extracting the same feature (e.g. IMC1). Table 2 shows the range of A_z values provided by each morphological feature alone, using the 2D and 3D active contour models. The ranges in Table 2 are for different methods of combining the features extracted from individual slices, i.e., mean, variance, minimum, and maximum. The most discriminatory feature in this study was the IMC1 feature (d=6, θ=0•, extracted from the upper disk-shaped region segmented by the 3D method) with an A_z value of 0.76.

When stepwise LDA was used to combine the features into a discriminant score in the 102 leave-one-case-out training subsets, an average of 6.09 and 7.98 feature were selected with the 2D and 3D segmentation methods, respectively. For the 2D segmentation method, the most frequently selected features were two IMC1 features, two IMC2 features, one DFE feature, and one width-to-height feature. For the 3D segmentation method, the most frequently selected features were two IMC1 features, two IMC2 features, one DFE feature, one ENT feature, one PSF feature and one width-to-height feature. Fig. 6 shows the test ROC curves obtained by the LDA using leave-one-case-out resampling for the 2D and 3D segmentation methods. The test A_z values for the 2D and 3D methods were 0.88±0.03 and 0.92±0.03, respectively, and the $A_z^{(0.9)}$ values were 0.51±0.10 and 0.65±0.09, respectively. The difference between the two test A_z values did not achieve statistical significance (p=0.16). Fig. 7 shows the distribution of the discriminant scores obtained from the 3D method for the malignant and benign cases.

In order to investigate the dependence of the classification accuracy on the initialization of the 3D active contour model, we scaled, rotated, and translated the initial 3D ellipsoid and repeated the steps of

active contour segmentation, feature extraction and classification for these modified initial ellipsoids. The classification accuracies for these experiments are presented in Table 3. None of the differences between the A_z values on Table 3 achieved statistical significance.

The ROC curves for radiologists' malignancy ratings are shown in Fig. 8. The computer and radiologist A_z values and $A_z^{(0.9)}$ values are compared in Table 4. The area A_z under the ROC curve for radiologists RAD1-RAD4 varied between 0.84±0.04 and 0.92±0.03, which are lower than or equal to that of the 3D computer classifier. The average A_z value, obtained by averaging the slope and intercept parameters (a and b in ROC analysis) of the individual ROC curves was 0.87. The difference between the A_z values of the individual radiologists and the computer classifiers (2D and 3D methods) did not reach statistical significance ($p>0.05$). The difference between the $A_z^{(0.9)}$ values of the individual radiologists and the classifier based on 2D segmentation also did not reach statistical significance ($p>0.05$) although the $A_z^{(0.9)}$ value of the computer classifier was consistently higher than those of all four radiologists. The $A_z^{(0.9)}$ values of the classifier based on 3D segmentation was again higher than the $A_z^{(0.9)}$ values of all four radiologists, and achieved statistical significance in three of the four ($p=0.02$, 0.04, and 0.008 for RAD1, RAD2, and RAD4, respectively).

IV. DISCUSSION

The computer classifier designed in this study to characterize breast masses on US volumes was able to discriminate between malignant and benign masses that were suspicious enough to warrant biopsy. From Fig. 7, it is observed that if an appropriate decision threshold was chosen for the discriminant scores of the classifier based on 3D segmentation, more than 45% (20/44) of biopsied benign masses could be correctly identified while no malignant masses were misclassified (at 100% sensitivity). Based on 2D segmentation the corresponding percentage of correctly identified benign masses was 36% (16/44).

Lesion segmentation is an important task in computerized lesion characterization. Segmentation of US images can be challenging because boundaries are not always conspicuous, due to the noise and contrast characteristics, and the speckled nature of US images. For breast US, an additional source of difficulty is the presence of posterior shadowing artifacts, a major source of which is the US attenuation due to the fibrous stroma caused by the tumor.[28] Previous research on segmentation of breast masses on US images includes work by Horsch et al.,[29] Xiao et al.,[30] and Madabhushi. et al.[31]. Their segmentation

methods were applied to 2D US images. In our study, we compared the classification accuracy when 2D and 3D active contour models were used for segmentation. The 2D model provided reasonable segmentation results for many of the masses. However, the 2D model does not take advantage of the image information in adjacent slices when a particular slice is being segmented. If the 2D active contour is misled on one slice, there is no interaction from adjacent slices to improve the segmentation. This is illustrated in Fig. 4, row 3. It can be observed that the 2D segmentation results on slices #45 and #47 are reasonable; however, part of the lesion is missed by the 2D active contour model on slice #46. Our 3D active contour model uses the smoothness of the segmented shape in the out-of-plane direction as an interaction term between adjacent slices. The 3D segmentation results, shown in row 4, are more consistent across slices. Figure 5 compares the segmented object using the 2D and 3D methods for the entire lesion, which was visible on a total of 10 slices. It is again observed that the lesion shape in the out-of-plane direction is smoother for the 3D method.

The texture features in this study were extracted from disk-shaped regions at the upper and lower margins of the mass on each slice. The total area of the two disk-shaped regions was equal to the area of the segmented mass. From Table 1, it is observed that a texture feature extracted from the upper disk-shaped region tended to be more discriminatory than the same feature extracted from the lower disk-shaped region. The maximum of the range of A_z values (the second number in each cell) was larger for the upper region in 11 of the 12 comparisons that can be made (6 texture features and 2 segmentation methods). The lower boundaries of many masses were difficult to perceive and hence difficult to automatically segment because of posterior shadowing. This may be contributing to the difference of discrimination ability between the features extracted from the upper and lower regions. Another possible factor may be the changes in the spatial and gray level resolutions in different regions of the US image as the distance from the US probe increases. Further work is underway to investigate the reasons for the apparent lower discrimination ability of the features extracted from the lower disk-shaped regions.

Although the disk-shaped region depends on mass segmentation, there can be a large overlap between the regions from the 2D and 3D segmentation results if the objects segmented by the two methods are not very different. From Table 1, it can be observed that the ranges of A_z values for 2D and 3D segmentation for each texture measure have a large overlap, especially for the first three feature measures in the table (IMC1, IMC2, and DFE), which were relatively more discriminatory. These were

also the features most frequently selected during classifier design. As mentioned in the Results Section, when the stepwise feature selection method was used for classifier design from 2D segmentation results, an average of 6.09 features were selected, where the average was computed over different cycles of the leave-one-out partitioning of the data set. Out of the 6 most frequently selected features, 5 were texture features and 1 was a morphological feature. The IMC1 feature was selected twice (at d=2, θ=0• and d=6, θ=90•), the IMC2 feature was selected twice (at d=2, θ=0• and d=6, θ=0•), and the DFE feature was selected once (at d=6, θ=0•). For 3D segmentation, out of the 8 most frequently selected features, 6 were texture features and 2 were morphological features. The IMC1 feature was selected twice (at d=2, θ=90• and d=4, θ=0•), the IMC2 feature was selected twice (at d=2, θ=0• and d=6, θ=0•), and the DFE feature was selected once (at d=6, θ=0•). Thus, out of 11 most frequently selected texture features (5 for 2D and 6 for 3D segmentation), 10 were IMC1, IMC2, or DFE features. The classification accuracy with the stepwise LDA for the 3D segmentation (A_z=0.92) was better than that for 2D segmentation (A_z=0.88). However, the difference did not achieve statistical significance (two-tailed p value = 0.16).

The active contour method requires an initial boundary to start iterating towards the optimal contour. In this study, the initial boundary was defined by a 3D ellipsoid that approximated the mass shape. The ellipsoid was placed in the volume by one of the radiologists (RAD1) using an interactive graphical user interface (GUI). The radiologist thus had to shift and scale a single object to define the initial contour. Although the error between the true and approximated shapes can be large when a single object is used for approximating the mass, this method was faster than other possible methods that would require initialization on each slice separately, and was therefore preferred. The robustness of the 3D segmentation method to active contour initialization was studied by translating, rotating, and scaling the 3D ellipsoid. There are many possibilities as to how these three operations (moving, rotating, and scaling) can be combined to modify the initial ellipsoid. In Table 3, the classification results are presented when these three operations are performed one at a time. Row 1 shows the A_z value when the original ellipsoid is used. The ellipsoid was scaled in rows 2-3, translated in rows 4-6, and rotated in row 7. For the magnitudes of scaling, translation and rotation studied in Table 3, the variation of the A_z value was within two standard deviations of A_z value provided by the LABROC program.[26] In a step towards automating the initialization of the contour, we are currently investigating methods for automatically determining an initial contour from a rectangular box containing the mass.

The comparison of the ROC curves by the radiologists and the computer indicated that the computer can be as effective as the radiologists in differentiating malignant and benign breast masses in this data set. In fact, the accuracy of the computer classifier using 3D segmentation was greater than three and equal to one of the radiologists, although the difference between the computer and the individual radiologists in terms of A_z did not achieve statistical significance. Furthermore, from Fig. 8, it is observed that the computer has a tendency to be better at high sensitivity. This was also confirmed by the statistically significant difference between the computer classifier (3D segmentation method) and three out of four radiologists when the comparison was based on the $A_z^{(0.9)}$ values. It should be noted that the purpose of our study was not to evaluate our US mass characterization method in a clinical setting, As noted in the Introduction and the Methods sections, the semi-automated 3D data acquisition system used in this study is still under investigation and was different from that in current clinical practice. The first difference is that, in our department, radiologists interactively perform hand-held US examination themselves, which may yield better image quality and may result in higher characterization accuracy. The second difference is that our study concentrated only on mass characterization of found lesions, whereas the actual detection of suspicious masses by US is a very important step in a clinical examination. These other aspects of comparing 3D US images to US images acquired with current clinical methods are subjects of future investigations.

In this study, the features were extracted from individual US slices and then combined into object-based features, as explained in Section 2.D. Although this method is found to provide effective features in this study, it may not have fully utilized the information available in the 3D data set. The potential improvement in classification accuracy by using truly 3D features, for example, texture features extracted from 3D SGLD matrices, needs to be investigated. Furthermore, in clinical practice, the decision about whether the mass is malignant or benign is made using both mammographic and US image information, as well as other pertinent patient information. A study is currently underway in our laboratory to design a classifier that combines computer-extracted features or scores from these two imaging modalities.

V. CONCLUSION

A computer segmentation and classification method has been developed for the task of characterization of breast masses on 3D US images. On a data set of 102 biopsy-proven masses the classifier achieved an A_z value of 0.92. The average A_z value of 4 experienced radiologists on the same

data set was 0.87. The computer classifier was more accurate than three and equal to one of the four radiologists participated in the study. However, the difference between the A_z values of the computer and the individual radiologists did not achieve statistical significance for this data set. At high sensitivity, the computer classifier was consistently more accurate than all 4 radiologists and achieve statistical significance ($p<0.05$) for the difference in $A_z^{(0.9)}$ from three of the four radiologists. The robustness of the iterative segmentation algorithm in terms of the initial contour provided to the algorithm was studied. The classification accuracy was found to depend on the initialization; however, the A_z value did not significantly deteriorate when the initial contour was scaled, rotated, or translated by a moderate amount. Future work includes verifying the results of this study by applying it to a larger and independent data set, expanding the feature space by designing truly 3D features, and combining the developed US characterization method with mammographic characterization methods. Observer performance study will also be performed to evaluate the effects of CAD on the characterization of breast masses by radiologists.

ACKNOWLEDGMENTS

This work was supported by USAMRMC grant DAMD17-01-1-0328 and by USPHS grants CA095153 and CA91713. The content of this publication does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned in the publication should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for providing the LABROC program.

REFERENCES

[1] L. W. Bassett, T. H. Liu, A. I. Giuliano, and R. H. Gold, "The prevalence of carcinoma in palpable vs impalpable, mammographically detected lesions," AJR 158, 688-689 (1992).

[2] H. Opie, N. C. Estes, W. R. Jewell, C. H. Chang, J. A. Thomas, and M. A. Estes, "Breast biopsy for nonpalpable lesions: a worthwhile endeavor?," American Surgeon 59, 490-493 (1993).

[3] G. Hermann, C. Janus, I. S. Schwartz, B. Krivisky, S. Bier, and J. G. Rabinowitz, "Nonpalpable breast lesions: Accuracy of prebiopsy mammographic diagnosis," Radiology 165, 323-326 (1987).

[4] F. M. Hall, J. M. Storella, D. Z. Silverstone, and G. Wyshak, "Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography," Radiology 167, 353-358 (1988).

[5] J. A. Baker, P. J. Kornguth, J. Y. Lo, and C. E. Floyd, "Artificial neural network: Improving the quality of breast biopsy recommendations," Radiology 198, 131-135 (1996).

[6] Z. M. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign masses on digitized mammograms," Acad. Radiol. 5, 155-168 (1998).

[7] H.-P. Chan, B. Sahiner, M. A. Helvie, N. Petrick, M. A. Roubidoux, T. E. Wilson, D. D. Adler, C. Paramagul, J. S. Newman, and S. S. Gopal, "Improvement of radiologists' characterization of mammographic masses by computer-aided diagnosis: an ROC study," Radiology 212, 817-827 (1999).

[8] A. T. Stavros, D. Thickman, C. L. Rapp, M. A. Dennis, S. H. Parker, and G. A. Sisney, "Solid breast nodules: Use of sonography to distinguish between malignant and benign lesions," Radiology 196, 123-134 (1995).

[9] P. Skaane and K. Engedal, "Analysis of sonographic features in differentiation of fibroadenoma and invasive ductal carcinoma," AJR 170, 109-114 (1998).

[10] K. J. W. Taylor, C. Merritt, C. Piccoli, R. Schmidt, G. Rouse, B. Fornage, E. Rubin, D. Georgian-Smith, F. Winsberg, B. Goldberg, et al., "Ultrasound as a complement to mammography and breast examination to characterize breast masses," Ultrasound Med. Biol. 28, 19-26 (2002).

[11] B. Sahiner, G. L. LeCarpentier, H.-P. Chan, M. A. Roubidoux, N. Petrick, M. M. Goodsitt, S. S. Gopal, and P. L. Carson, "Computerized characterization of breast masses using three-dimensional ultrasound images," Proc. SPIE 3338, 301-312 (1998).

[12] D. R. Chen, R. F. Chang, and Y. L. Huang, "Computer-aided diagnosis applied to US of solid breast nodules by using neural networks," Radiology 213, 407-412 (1999).

[13] K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny, "Computerized diagnosis of breast lesions on ultrasound," Med. Phys. 29, 157-164 (2002).

[14] C. M. Chen, Y. H. Chou, K. C. Han, G. S. Hung, C. M. Tiu, H. J. Chiou, and S. Y. Chiou, "Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks," Radiology 226, 504-514 (2003).

[15] D. B. Downey, A. Fenster, and J. C. Williams, "Clinical utility of three-dimensional US," Radiographics 20, 559-571 (2000).

[16] G. L. LeCarpentier, P. B. Tridandapani, J. B. Fowlkes, M. A. Roubidoux, A. P. Moskalik, and P. L. Carson, "Utility of 3D ultrasound in discriminating and detection of breast cancer," RSNA EJ http://ej.rsna.org/ej3/0103-99.fin/titlepage.html, (1999).

[17] P. T. Bhatti, G. L. LeCarpentier, M. A. Roubidoux, J. B. Fowlkes, M. A. Helvie, and P. L. Carson, "Discrimination of sonographically detected breast masses using frequency shift color Doppler imaging in combination with age and gray scale criteria," Journal of Ultrasound in Medicine 20, 343-350 (2001).

[18] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," Int. J. Comput. Vision 1, 321-331 (1987).

[19] L. D. Cohen, "On active contour models and balloons," CVGIP: Img. Underst. 53, 211-218 (1991).

[20] D. J. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation," CVGIP: Img. Underst. 55, 14-26 (1992).

[21] B. S. Garra, B. H. Krasner, S. C. Horri, S. Ascher, S. K. Mun, and R. K. Zeman, "Improving the distinction between benign and malignant breast lesions: The value of sonographic texture analysis," Ultrasonic Imaging 15, 267-285 (1993).

[22] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," IEEE Trans. Sys. Man. and Cybern. SMC-3, 610-621 (1973).

[23] Y. Zheng, G. JF, and G. JJ, "Reduction of breast biopsies with a modified self-organizing map," IEEE Trans. Neural Net. 8, 1386-1396 (1997).

[24] P. A. Lachenbruch, *Discriminant Analysis*, (Hafner Press, New York, 1975).

[25] N. R. Draper, *Applied regression analysis*, (Wiley, New York, 1998).

[26] C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," Stat. Med. 17, 1033-1053 (1998).

[27] Y. Jiang, C. E. Metz, and R. M. Nishikawa, "A receiver operating characteristic partial area index for highly sensitive diagnostic tests," Radiology 201, 745-750 (1996).

[28] B. Mesurolle, M. Ariche-Cohen, F. Mignon, J. Guinebretiere, A. Tardivon, and P. Goumot, "Small focal areas of acoustic shadowing in the breast," Journal of Clinical Ultrasound 88-97 (2002).

[29] K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny, "Automatic segmentation of breast lesions on ultrasound," Med. Phys. 28, 1652-1659 (2001).

[30] G. Xiao, M. Brady, J. Noble, and Z. Yongyue, "Segmentation of ultrasound B-mode images with intensity inhomogeneity correction," IEEE Trans. Med. Img. 21, 48-57 (2002).

[31] Madabhushi A and M. DN, "Combining low, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions," IEEE Trans. Med. Img. 22, 155-169 (2003).

Texture Feature	3D Segmentation		2D Segmentation	
	Upper	Lower	Upper	Lower
IMC1	0.65-0.76	0.58-0.68	0.65-0.73	0.58-0.66
IMC2	0.65-0.74	0.58-0.66	0.65-0.73	0.61-0.68
DFE	0.61-0.70	0.63-0.68	0.58-0.69	0.64-0.70
ENT	0.58-0.64	0.53-0.60	0.63-0.67	0.57-0.63
ENE	0.56-0.61	0.50-0.58	0.54-0.61	0.50-0.54
SME	0.52-0.58	0.51-0.55	0.54-0.62	0.51-0.57

Table 1: The range of A_z values for different texture features extracted from the lower and upper disk-shaped regions using the 3D and 2D segmentation methods. For each particular texture feature (e.g., IMC1 feature at pixel-pair distance d=2, and direction θ=0•), the feature values from all the slices containing the segmented mass were averaged before computing the A_z value. The range indicates the minimum-maximum A_z values for a particular feature among the parameters d=2, 4, 6 and θ=0•, 90•.

Morphological Feature	3D Segmentation	2D Segmentation
Width-to-height	0.60-0.71	0.60-0.69
PSF	0.51-0.66	053-0.61

Table 2: The range of A_z values for the width-to-height feature and posterior shadowing feature (PSF) extracted using the 3D and 2D segmentation methods. The range indicates the minimum-maximum A_z values among the mean, variance, minimum, and maximum of each feature extracted from each slice containing the segmented mass.

Scale	Rotation (degrees)	x-translation (pixels)	y-translation (pixels)	A_z
1	0	0	0	0.92±0.03
1.3	0	0	0	0.87±0.04
0.8	0	0	0	0.89±0.03
1	0	10	10	0.89±0.03
1	0	10	-10	0.85±0.04
1	0	-10	10	0.88±0.03
1	0	-10	-10	0.87±0.04
1	15	0	0	0.92±0.03

Table 3: The dependence of the computer classification accuracy on the variation of the initial contour. The effects of three transformation parameters, namely, scaling, translation and rotation of the initial ellipsoid, was investigated by moving the initial ellipsoid using one of these three parameters at a time. A translation by ±10 pixels in the image plane corresponded to approximately ±1 mm.

	A_z	$A_z^{(0.9)}$
Computer classifier, 2-D segmentation	0.88±0.03	0.51±0.10
Computer classifier, 3-D segmentation	0.92±0.03	0.65±0.09
RAD1	0.84±0.04	0.41±0.10*
RAD2	0.86±0.04	0.37±0.11*
RAD3	0.92±0.03	0.44±0.14
RAD4	0.84±0.04	0.28±0.11*

Table 4: The area under the ROC curve (Az), and the area under the ROC curve above a sensitivity threshold of 0.9 ($A_z^{(0.9)}$) for the computer classifier using the 2-D and 3-D active contour segmentation results, and the four radiologists. The radiologists' results that are significantly ($p<0.05$) different from the 3-D computer results are noted with an asterisk.

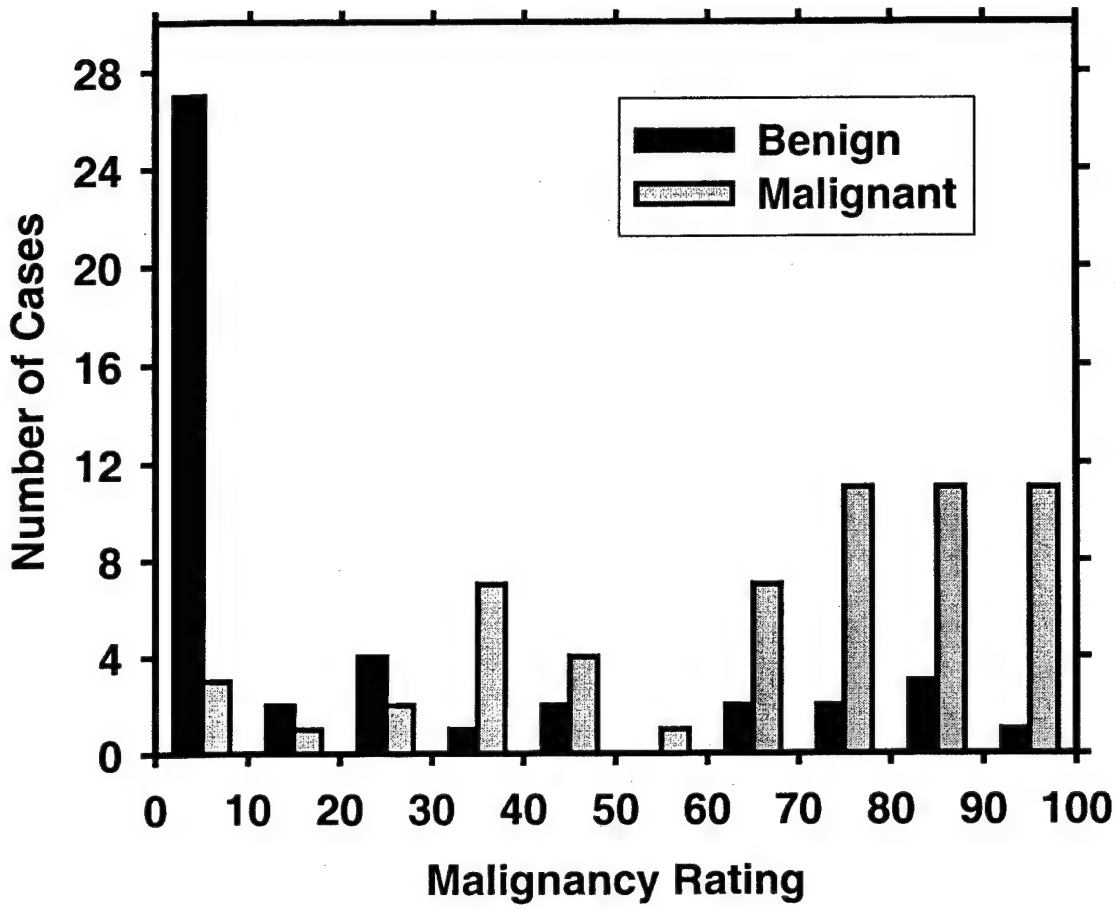

Figure 1. The distribution of the malignancy rating of the masses in our data set based on the appearance on US images, by an experienced radiologist. 1: Very likely benign, 100: Very likely malignant.

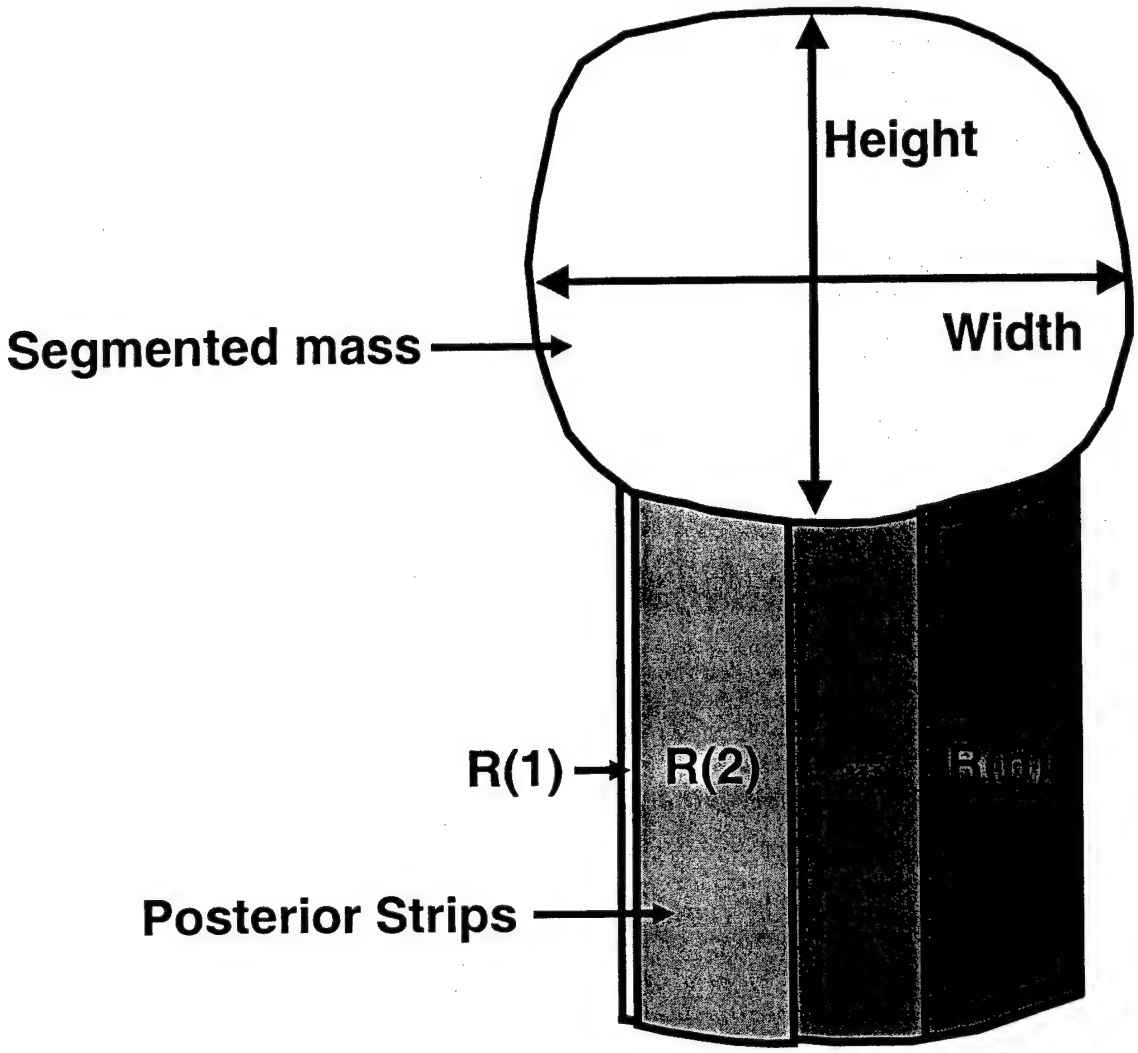

Figure 2. The definition of the width-to-height and PSF features. The width-to-height feature was defined as the ratio of the widest cross-section of the segmented mass shape in the image plane to the tallest cross-section. The PSF feature was defined by first finding the average gray value in the posterior strips $\overline{R(i)}$, $i = 1,\ldots,n$, then finding the minimum of $\overline{R(i)}$, and finally by normalizing this value by the average gray value within the segmented mass.

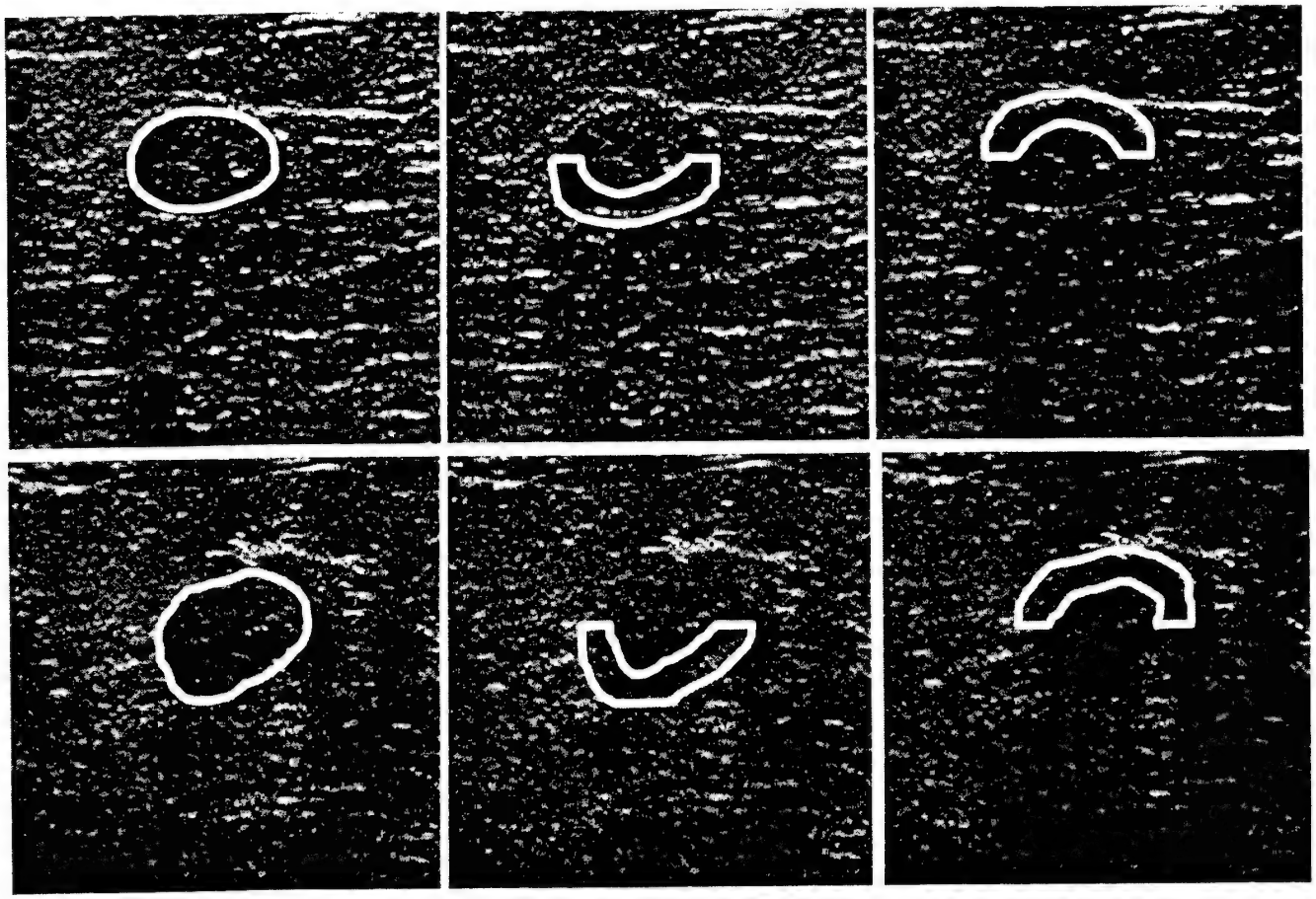

Figure 3. Left column: The segmented object for a malignant mass (upper row) and a benign mass (lower row). Middle and right rows: The lower and upper disk-shaped regions from which texture features were extracted.

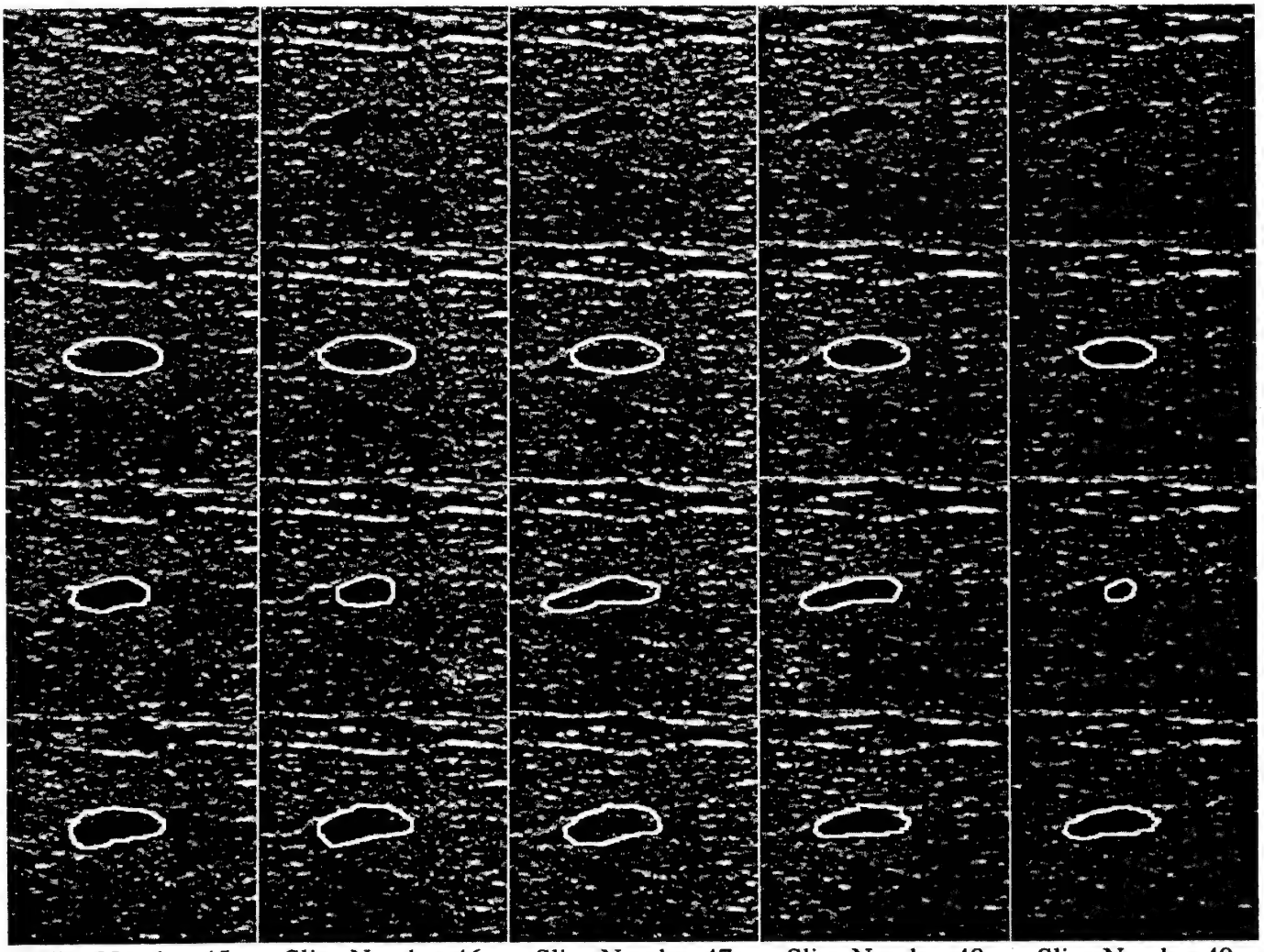

Figure 4. Row 1: Five original slices of a breast mass that was visible on a total of ten US slices; Row 2: The cross section of the initial 3D ellipsoid at each slice; Row 3: The result of the 2D active contour segmentation method; Row 4: The result of the 3D active contour segmentation method. Note that the 2D segmentation method misses part of the mass on slice 46. The 3D segmentation method, apparently using the information from slices 45 and 47, is able to provide better segmentation on slice 46.

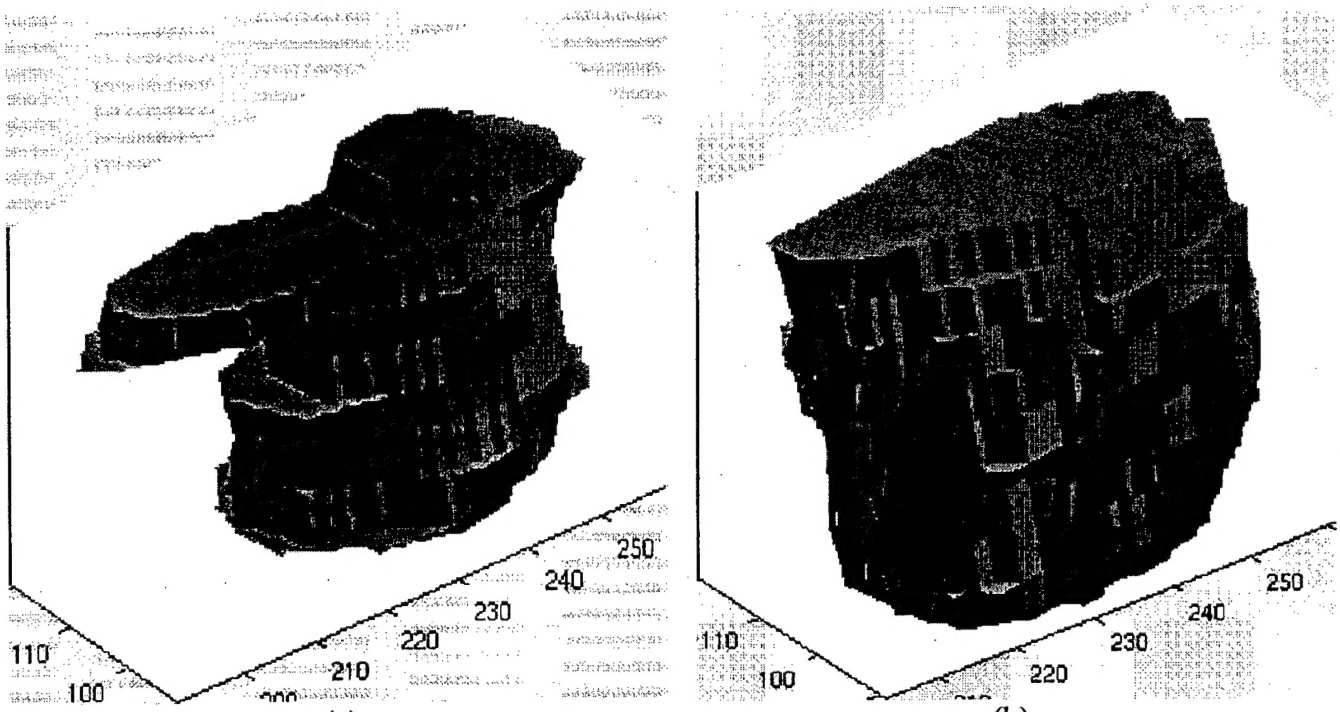

Figure 5. 3D rendering of the segmented object for the mass shown in Fig. 4. (a) 2D active contour segmentation; (b) 3D active contour segmentation.

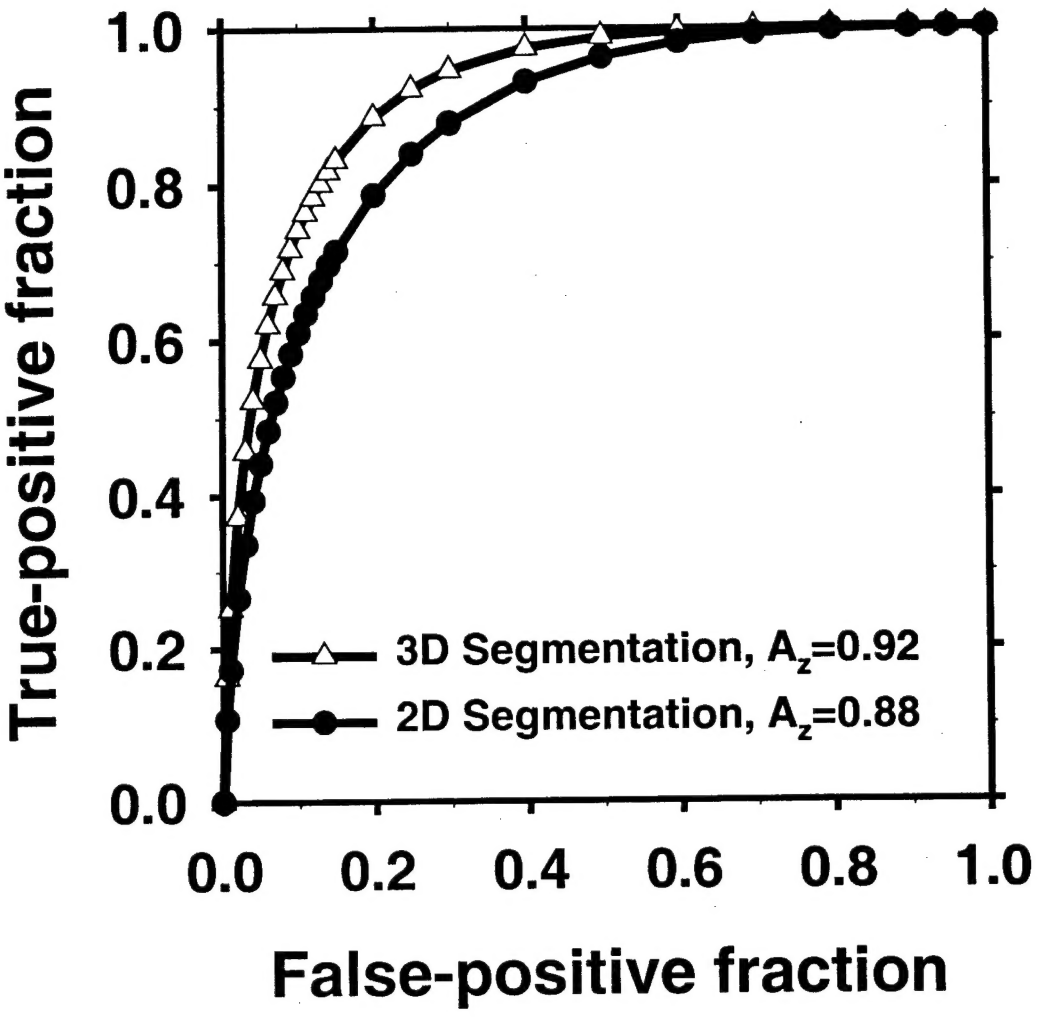

Figure 6. The test ROC curves obtained by the classifiers that were based on features extracted from the 2D (A_z=0.88) and 3D (A_z=0.92) active contour models. The difference between the two A_z values did not achieve statistical significance (p=0.16).

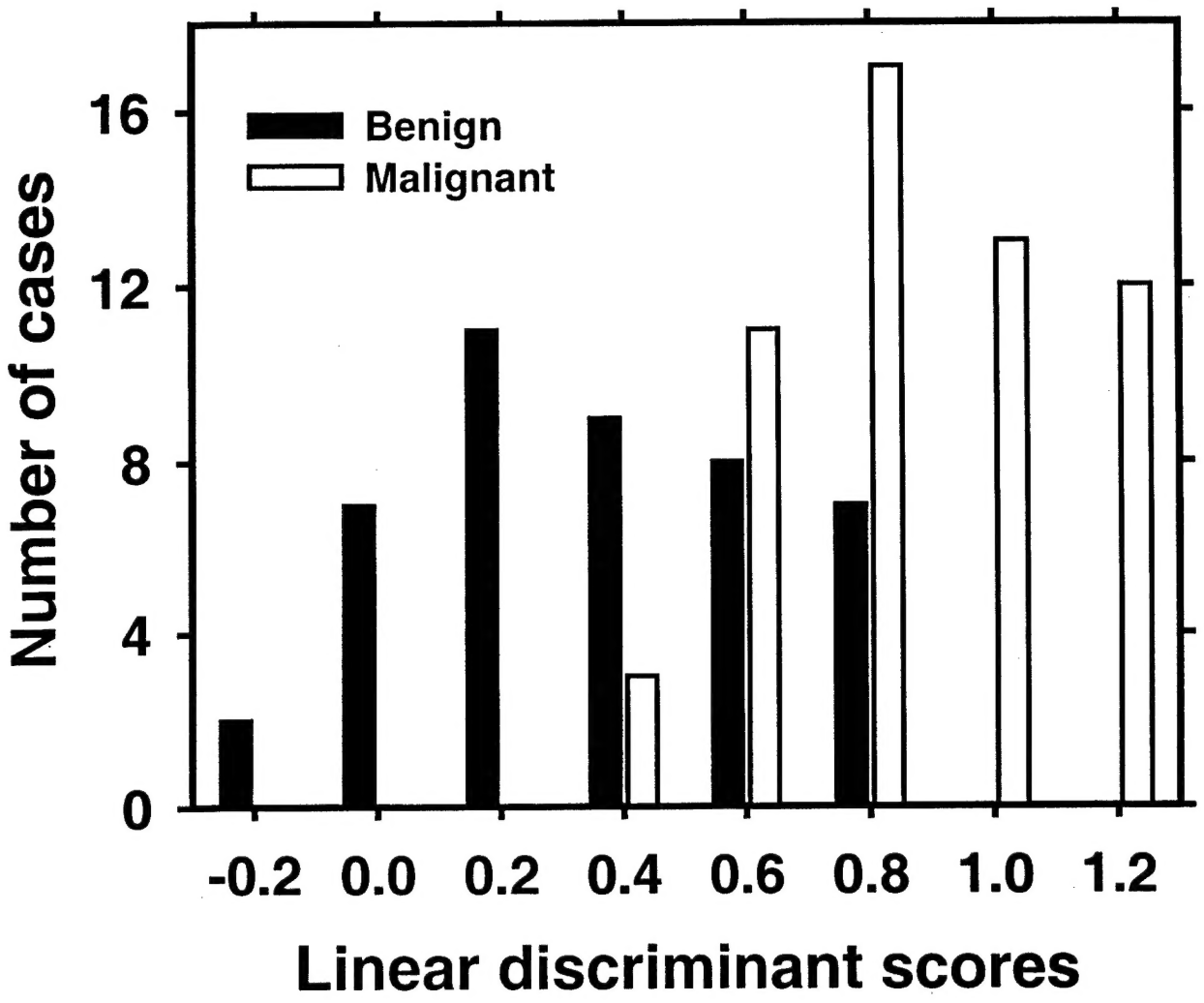

Figure 7. The distribution of the test discriminant scores for the classifier that was based on 3D active contour segmentation. By choosing an appropriate decision threshold on these scores, (e.g., decision threshold=0.3) more than 45% (20/44) of biopsied benign masses could be correctly identified while no malignant masses were misclassified.

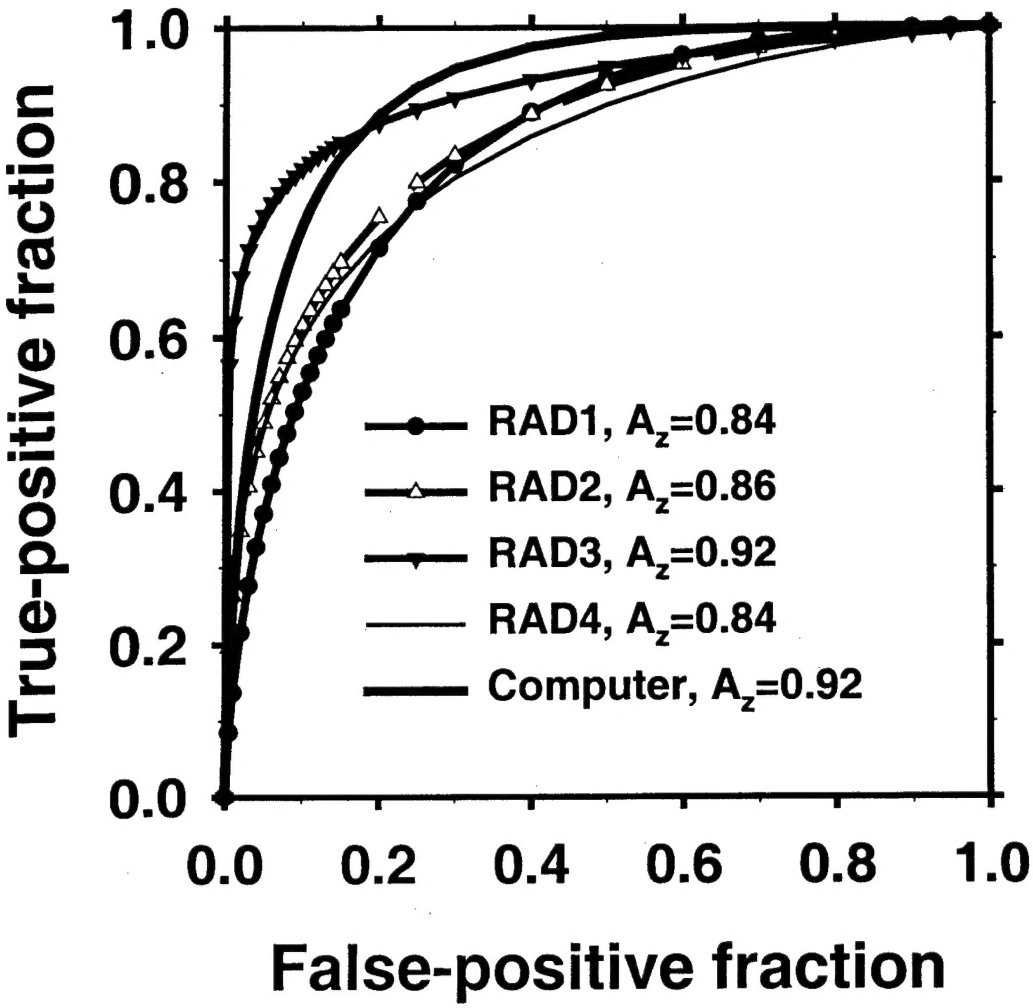

Figure 8. ROC curves for the computer and for the four radiologists who participated in the malignancy rating experiment. The difference between the computer's A_z value and that of any of the four radiologists did not achieve statistical significance. However, the computer classifier was significantly ($p<0.05$) more accurate than three of the four radiologists at high sensitivity (TPF>0.9).